ANALYSIS OF BIOMARKERS FOR COMPLEX HUMAN DISEASES

Morad Ansari

Thesis presented for degree of Doctor of Philosophy

The University of Edinburgh

2009

Declaration

I declare that this thesis was composed by me. The contributions of others to the work are clearly indicated. This work has not been submitted for any other degree or professional qualification.

Morad Ansari
April 2008

Acknowledgments

First and foremost my thanks go to my supervisor, *Alan Wright*, for giving me the opportunity to work on this exciting project. Your vast knowledge has been an inspiration to me and thank you for your support, help and guidance through the course of this project. Secondly, I would like to thank my second supervisor, *Nick Hastie*, for his interest in the study and for many helpful discussions.

I am very grateful to *Caroline Hayward* and *Veronique Vitart*; without your expert help and support this project would simply have not been possible. I would also like to thank *Dafni Vlachantoni*, firstly for being a great friend and secondly for her valuable help with the thesis. Also thanks to the other post-docs in the group, *Xinhua Shu* and *Kevin Chalmers* for help with various parts of the project. Thanks are also due to *Susan Campbell* and *Joanne Morgan* for assisting me with the CFH ELISA.

At the MRC Human Genetics Unit, my thanks go to *Peter Teague* for help with the statistical analysis of data, *Phillip Gautier* for the bioinformatics analysis, *Rebecca Barnetson* for help with the quantitative PCR and *Craig Nicol* for photographic support.

At the University of Edinburgh, my thanks go to *Sara Knott* for advice on statistical analyses and to *Sander Henzing*, *James Creanor* and *Logan MacKay* for teaching me about mass spectrometry and proteomic techniques.

I would like to express my gratitude to *Igor Rudan* and his colleagues at the University of Zagreb, for providing the Croatian samples. I would also like to thank our other collaborators, *Steven Younkin* for providing the Aβ assay, *Ben Oostra* for providing the Dutch samples and *Takeshi Iwata* for the Japanese samples.

During the past few years I met some great people, so here is to everyone on E4 especially *Kirsty*, *Susan* and *Richard* and to my other friends *David*, *Jonathan*, *James* and *Jayne*, for all the good times in Edinburgh.

I dedicate this work to my family. Firstly to *Lisa*, thank you for your patience especially during the past few months and for making life worth living. I would not have been able to get through this without you. Secondly, to *Narges*, *Mina* and *Omid* for believing in me and for always encouraging me to achieve more.

I would like to end with a quote from the Persian philosopher, Avicenna,

“The human intellect at birth is rather like a ‘tabula rasa’, a pure potentiality that is actualised through education and comes to know" and is developed through a "syllogistic method of reasoning; observations lead to prepositional statements, which when compounded lead to further abstract concepts."

Dedication

I dedicate this to my son, my life and joy, Julian.

TABLE OF CONTENTS

Abstract i
List of abbreviations iii
List of Figures x
List of Tables xiii

1 - CHAPTER 1: INTRODUCTION 1
1.1 Mapping genes for common, complex human diseases 2
1.1.1 Genetic architecture of complex traits 5
1.1.2 Lessons from association/linkage mapping of QTL 6
1.1.3 Mendelian randomisation 7
1.1.4 Choice of population 8
1.1.4.1 Gene mapping in isolated populations 9
1.1.5 The Croatia project 12
1.2 Proteomics and the human plasma 14
1.2.1 Analysis of plasma proteins 15
1.2.1.1 Enzyme-linked immunosorbent assay 16
1.2.1.2 Mass spectrometry 17
1.2.1.2.1 Top-down' approach 22
1.2.1.2.2 Bottom-up' approach 23
1.3 Alzheimer disease 25
1.3.1 Genetics of Alzheimer disease 30
1.3.2 Biochemistry and genetics of amyloid-β 31
1.3.3 Rodent models of Alzheimer disease 34
1.4 Biochemistry and genetics of APOE 37
1.4.1 APOE and dementia 38
1.5 Age-related macular degeneration 39
1.5.1 Genetics of AMD 43
1.5.2 AMD and the complement pathway 44
1.5.3 AMD association with chromosome 10q 50
1.5.4 Other single gene associations with AMD 51
1.5.5 Meta-analysis of established genetic associations with AMD 52
1.5.6 Animal models of AMD 53
1.6 Aims 55

2 – CHAPTER 2: MATERIALS AND METHODS 59
2.1 Introduction 60
2.2 Sample collection and description 60
2.3 Molecular genetic techniques 63
2.3.1 APOE genotyping 63
2.3.1.1 APOE genotyping by RFLP 65
2.3.1.2 Sequencing of *APOE*E2* carriers 66
2.3.1.3 Experimental controls 67
2.3.1.4 Genotyping replications 67
2.3.2 Genotyping of STR markers 67
2.3.3 Genotyping of single nucleotide polymorphisms 67
2.3.3.1 Whole genome SNP genotyping 68

2.3.3.2 SNP genotyping by TaqMan® 5′nuclease assay....68
2.3.3.3 SNP genotyping by direct sequencing....69
2.3.4 Genotyping of the *CFHR3/CFHR1* deletion....70
2.3.4.1 Theory of quantitative polymerase chain reaction....70
2.3.4.2 Sequence analysis at the *CFH* locus....71
2.3.4.3 Multiplex PCR of *CFHR3/CFHR1* deletion....72
2.3.4.4 Quantitative PCR of *CFHR3/CFHR1* deletion....74
2.3.4.4.1 Designing the qPCR primers and probes....74
2.3.4.5 Optimisation of primer and probe concentrations....74
2.3.4.5.1 Experimental approach....75
2.4 Biochemical techniques....75
2.4.1 Commercial assay of amyloid-β in plasma....75
2.4.2 Optimised assay of amyloid-β in plasma....76
2.4.3 Assay of complement factor H in plasma....77
2.4.3.1 Plasma extraction from blood....78
2.4.3.2 CFH assay optimisation....78
2.4.3.3 Protocol....80
2.4.4 Protein separation by electrophoresis....81
2.4.4.1 Total protein quantitation....81
2.4.4.2 Protein separation using Laemmli gels....82
2.4.4.3 Protein separation using pre-cast gels....82
2.4.5 Semi-dry transfer of proteins to membranes....83
2.4.6 Western blotting....83
2.4.7 SDS-PAGE gel staining....84
2.4.7.1 SimplyBlue™ staining....84
2.4.7.2 Silver staining....85
2.4.8 Proteomic and mass spectrometry methods....85
2.4.8.1 Immunodepletion of high abundance plasma proteins....86
2.4.8.1.1 Background....86
2.4.8.1.2 Protocol....86
2.4.8.2 Hydrophobic interaction chromatography....87
2.4.8.2.1 Background....87
2.4.8.2.2 Protocol....87
2.4.8.3 Pretreatment with ZipTip™....88
2.4.8.4 Ultrafiltration of plasma....88
2.4.8.4.1 Background....88
2.4.8.4.2 Protocol....89
2.4.8.5 MALDI-TOF MS analysis....89
2.4.8.6 Capillary electrophoresis coupled with MS....89
2.4.8.6.1 CE analysis of single protein sample....90
2.4.8.6.2 CE troubleshooting....91
2.5 Statistical analyses....91
2.5.1 Predictive power calculation....91
2.5.2 Empirical analysis of power using plasma lipids....92
2.5.3 Normalising inter-experimental variation in ELISA of CFH and Aβ....93
2.5.4 Detection of outliers in assays of CFH and Aβ....93
2.5.5 Testing for normality....94
2.5.6 Quantile normalisation....95
2.5.7 Linear regression analyses....95
2.5.7.1 Multiple regression....95
2.5.7.2 Stepwise linear regression analysis of markers....98
2.5.7.2.1 Detecting collinearity among predictors....98

2.5.7.3 Multi-locus association analysis....99
2.5.7.4 The linear mixed model....99
2.5.7.5 Estimates of heritability....100
2.5.7.6 Genome-wide association analysis....100
2.5.7.6.1 Correction for multiple testing....101
2.5.8 Genome-wide linkage analysis....102
2.5.9 Estimation of Hardy-Weinberg equilibrium....103
2.5.10 Analysis of linkage disequilibrium....103
2.5.11 Bioinformatic resources....105
2.6 List of reagents....105

3 - CHAPTER 3: ANALYSIS OF POWER 109
3.1 Introduction....110
3.2 Results....111
3.2.1 *APOE* genotyping in Vis sample set....111
3.2.2 Empirical analysis of power using plasma lipids....113
3.2.3 Predictive analysis of power....116
3.2.4 Summary and discussion....118

4 - CHAPTER 4: BIOCHEMICAL AND LINKAGE ANALYSES OF PLASMA AMYLOID-β 126
4.1 Introduction....127
4.2 Results....129
4.2.1 Commercial assay of Aβ40 and Aβ42 in plasma....129
4.2.2 Optimised assay of Aβ40 and Aβ42 in plasma....132
4.2.3 Analysis of plasma Aβ covariates in Vis....135
4.2.4 Plasma Aβ heritability in Vis....137
4.2.5 A genome-wide linkage scan of plasma Aβ in the Vis population....137
4.2.6 Summary and discussion....138

5 - CHAPTER 5: GENOME-WIDE ASSOCIATION ANALYSIS OF PLASMA Aβ 148
5.1 Introduction....149
5.2 Results....150
5.2.1 Genome-wide association analysis of plasma Aβ in Vis....150
5.2.1.1 QTL association analysis of plasma Aβ40....150
5.2.1.2 QTL association analysis of plasma Aβ42....157
5.2.1.3 QTL association analysis of the ratio of Aβ42:Aβ40 in plasma....162
5.3 Summary and discussion....168

6 - CHAPTER 6: BIOCHEMICAL AND LINKAGE ANALYSES OF PLASMA COMPLEMENT FACTOR H 176
6.1 Introduction....177
6.2 Testing the purity of CFH standard and specificity of OX23 antibody....179
6.3 CFH assay optimisation....179
6.3.1 Optimisation of antibody and standard concentrations....180
6.3.2 Optimisation of sample dilutions....180
6.3.3 Optimisation of incubation time with TMB chromogen....180
6.3.4 Analysis of intra-experimental variation....182
6.3.5 Analysis of inter-experimental variation....182
6.3.6 Analysis of plasma CFH concentration after sample freeze- thawing....184
6.3.7 Analysis of pipetting techniques in CFH assay....184
6.4 Plasma CFH measurements....186
6.4.1 Analysis of plasma CFH in the island of Vis....186
6.4.1.1 Analysis of covariates....186

6.4.2 Plasma CFH distribution in the isolated Rucphen (Dutch) community ..189
6.4.2.1 Analysis of covariates ..190
6.4.3 Analysis of plasma CFH in the Scottish AMD series ..190
6.4.3.1 Analysis of plasma CFH covariates in the Scottish AMD cases ..190
6.4.3.2 Analysis of plasma CFH covariates in the Scottish AMD control sample ..192
6.4.3.3 Comparison of plasma CFH concentrations between Scottish AMD cases and controls ..192
6.4.4 Analysis of plasma CFH in a Japanese AMD series ..194
6.4.5 Comparison of plasma CFH in general populations of Vis, Rucphen, Scotland and Japan ..194
6.5 Genome-wide linkage analysis of plasma CFH ..196
6.6 Summary and discussion ..197

7 - CHAPTER 7: GENETIC ASSOCIATION ANALYSIS OF PLASMA CFH 205
7.1 Introduction ..206
7.2 Results ..207
7.2.1 Genotyping of the *CFHR3/CFHR1* deletion ..207
7.2.1.1 Results of *CFHR3/CFHR1* deletion genotyping ..207
7.2.2 QTL Association analysis of CFH in Vis ..209
7.2.2.1 *CFH* SNP genotyping ..209
7.2.2.2 Genome-wide QTL association analysis of CFH ..211
7.2.2.3 Analysis of linkage disequilibrium across the *CFH* locus ..217
7.2.2.4 Identification of a multi-marker model for plasma CFH concentration ..219
7.2.2.5 Haplotype-based association analysis of CFH ..222
7.2.3 QTL analysis of plasma CFH in the Dutch sample set ..225
7.2.3.1 Genotyping of markers ..225
7.2.3.2 Single marker association analysis of plasma CFH ..225
7.2.4 Analysis of *CFH* association with AMD ..228
7.2.4.1 Analysis of haplotype association with AMD ..235
7.2.5 AMD association analysis in the Japanese series ..237
7.3 Summary and discussion ..239

8 - CHAPTER 8: ANALYSIS OF PROTEINS BY MASS SPECTROMETRY 252
8.1 Introduction ..253
8.2 Results ..254
8.2.1 Immunodepletion of high abundance plasma proteins ..254
8.2.2 MALDI-TOF mass spectrometry ..256
8.2.2.1 MALDI-TOF MS of the small mass range ..257
8.2.2.2 MALDI-TOF MS of the medium mass range ..262
8.2.2.3 MALDI-TOF MS of the large mass range ..265
8.2.3 Capillary electrophoresis ..267
8.2.3.1 Optimisation of peptide concentration ..267
8.2.3.2 Optimisation of injection times ..270
8.2.3.3 CE analysis of single protein samples ..270
8.2.3.4 CE coupled with FT-ICR MS ..271
8.2.4 Ion exchange chromatography ..272
8.3 Summary and discussion ..274

9 - CHAPTER 9: DISCUSSION 283
10 - CHAPTER 10: REFERENCES 303

ABSTRACT

The aims of this study were to analyse known and potential biomarkers of common and genetically complex human disorders and to identify genetic and environmental variation associated with plasma biomarker concentrations. Two groups of protein biomarkers were analysed. First, plasma complement factor H (CFH) was selected as a potential biomarker for age-related macular degeneration (AMD), since common variants in the *CFH* gene show strong association with this disorder. Secondly, two isoforms of amyloid-β (Aβ40 and Aβ42) were selected as biomarkers for Alzheimer disease (AD) since Aβ deposits are major constituents of the amyloid plaques characteristic of this disorder.

Physiological and anthropometric measurements and samples of human plasma and genomic DNA were collected from a population sample of 1,021 individuals from the Croatian island of Vis. Quantitative determination of plasma CFH, Aβ40 and Aβ42 concentrations was performed using enzyme-linked immunosorbent assays. Heritabilities and significant covariate effects were estimated for each trait in the Croatian data set. Genome-wide linkage and association analyses were conducted for the biomarker traits.

A novel finding was the genome-wide significant association between plasma CFH and several polymorphisms close to and within the *CFH* gene. The strongest association was with an intronic SNP within *CFH*, which explained 28% of the total trait variance ($P < 10^{-50}$). The association was also replicated in a Dutch sample set. A SNP haplotype was identified which accounted for a higher proportion of the phenotypic variance. Conditional haplotype analysis showed that the effect of this haplotype on plasma CFH concentration was independent of the *CFH* Y402H variant, and significantly stronger than a deletion of the adjacent *CFHR3*/*CFHR1* genes, which was already known to affect AMD susceptibility. Genetic analysis of 382 AMD cases and 201 controls was consistent with the *CFH* Y402H variant being the strongest AMD susceptibility locus. Variation in plasma CFH concentration was found to explain up to 1.8% of the variation in susceptibility to AMD with an odds ratio of 2.1 (95% C.I. 1.3-3.4, $P = 0.003$). SNPs that were strongly associated with plasma CFH concentration also influenced AMD susceptibility ($P < 0.05$) independently of the *CFH* Y402H polymorphism. Functional analysis of genomic regions associated with plasma CFH is needed to identify the causal variants.

Associations were observed between plasma Aβ40 concentration and several novel candidate loci, spanning regions of approximately 0.2 Mb, on chromosomes 9 and X. Similarly, novel associations with plasma Aβ42 were found in several regions, each spanning 0.2-0.4 Mb, on chromosomes 2, 5, 9, 15 and 20. The proportion of the phenotypic variance in plasma Aβ42 explained by these putative associations ranged between 1.8 and 2.8%. However, none of the associated SNPs was significant after correction for multiple testing, therefore replication is required.

Finally, attempts were made to identify and quantitate new protein biomarkers of disease in human plasma using mass spectrometry. Development and optimisation of techniques was initially undertaken to deplete high-abundance plasma proteins and improve signal:noise ratio. This allowed the assessment of downstream proteomic approaches including MALDI-TOF mass spectrometry (MS), capillary electrophoresis (CE) and ion exchange chromatography (IEC), each with the potential for large-scale quantitation of plasma proteins. Although the analysis of single protein analytes, using CE and IEC proved promising, the results highlighted the difficulty associated with MALDI-TOF and protein ionisation techniques in analysing complex mixtures such as plasma.

LIST OF ABBREVIATIONS

3D	Three dimensional
α-cyano	α-cyano-4-hydroxycinnamic acid
A	Adenine (DNA)
A	Absorbance
Aβ	Amyloid-β
ABC	ATP-binding cassette
ACN.	Acetonitrile
AD	Alzheimer disease
AEBSF	4-(2-Aminoethyl) benzenesulfonyl fluoride
AICD	APP intracellular domain
Ala	Alanine (amino acid)
AMD	Age-related macular degeneration
APP	Amyloid precursor protein
AX	Anion exchange chromatography
BLAST	Basic Local Alignment Search Tool
BMI	Body mass index
bp	Base pairs (DNA)
BrM	Bruch's membrane
BSA	Bovine serum albumin
C	Cytosine (DNA)
C	Chromatography
C.I.	Confidence interval
C2	Complement component 2
C3	Complement component 3
CDCV	Common disease-common variant
cDNA	Complementary DNA
CE	Capillary electrophoresis
CEPH	Centre d'Etude du Polymorphisme Humain
CFH	Complement factor H
CFHR	Complement factor H-related
Cfo	*Clostridium formicoaceticum*
cM	Centimorgan
cm	Centimetre
cm^2	Square centimetre

CNV	Choroidal neovascularisation
CR1	Complement component receptor 1
CRP	C-reactive protein
CSF	Cerebrospinal fluid
C_t	Threshold cycle
Cu	Cupper metal ion affinity chromatography
CV	Coefficient of variation
CX	Cation exchange chromatography
D	Deleted allele (*CFHR3*/*CFHR1* deletion)
d.f.	Degree of freedom
Da	Daltons
DAF	Decay accelerating factor
dATP	Deoxyadenosine triphosphate
dCTP	Deoxycytidine triphosphate
DGC	Dystrophin-glycoprotein complex
dGTP	Deoxyguanosine triphosphate
dm^3	cubic decimetre
DNA	Deoxyribonucleic acid
dNTP	Deoxyribonucleotide triphosphate
DTT	Dithiothreitol
dTTP	Deoxythymidine triphosphate
E	Efficiency
e.g.	For example (*exempli gratia*)
ECL	Enhanced chemiluminescence
EDTA	Ethylenediaminetetraacetic acid
EGF	Epidermal growth factor
ELISA	Enzyme-linked immunosorbent assay
EOF	Electro-osmotic flow
ESD	Extreme studentised deviate
ESI	Electrospray ionisation
et al.	And others (*et alii/aliae*)
F	Fraction
FA	Formic acid
FAD	Familial Alzheimer disease
FDR	False discovery rate

FHL	Factor H-like
fmol	femtomoles
FN3	Fibronectin type 3
FT-ICR	Fourier transform ion cyclotron resonance
g	G-force
G	Guanine (DNA)
g	Grams
GA	Geographic atrophy
GLM	General linear model
GRR	Genotype relative risk
GTP	Guanosine triphosphate
H	Histidine (amino acid)
h	Hour
h^2	Narrow-sense heritability
HBB	Beta haemoglobin gene
HDL	High-density lipoprotein
HGU	Human Genetics Unit
Hha	*Haemophilus haemolyticus*
HIC	High interaction chromatography
HPLC	High-performance liquid chromatography
HRP	Horseradish peroxidase
HST	Haplotype specific tests
HWE	Hardy-Weinberg equilibrium
i.e.	In other words (*illud est*)
IBD	Identity by descent
IBS	Identity by state
ICAT	Isotope coded affinity tags
ID	Identification
IDE	Insulin degrading enzyme
IEC	Ion exchange chromatography
Ig	Immunoglobulin
iTRAQ™	Isobaric tags for relative and absolute quantitation
k	Kilo
kb	Kilo basepairs (DNA)
kDa	Kilo Daltons

KI	Knock-in
KO	Knock-out
kV	Kilovolts
LC	Liquid chromatography
LD	Linkage disequilibrium
LDL	Low-density lipoprotein
LDS	Lithium dodecyl sulfate
LOAD	Late-onset Alzheimer disease
LOD	Likelihood of odds
M	Molar
m/z	Mass to charge ratio
mA	Milliampere
MAC	Membrane attack complex
MACPE	Membrane attack complex/perforin
MAF	Minor allele frequency
MALDI	Matrix-assisted laser desorption/ionisation
MB	Magnetic bead
Mb	Mega basepairs
MBS	Magnetic bead separator
MCP	Membrane co-factor protein
mg	Milligram
μg	Micrograms
MGB	Minor groove binder
MHC	Major histocompatibility complex
MHF	Minor haplotype frequency
min	minute
μJ	Microjoule
ml	Millilitre
μl	Microlitre
mm	Millimetre
μm	Micrometre
mM	Millimolar
μM	Micromolar
mm^3	cubic milimetre
mol	Moles

MOPS	3-(N-morpholino) propane sulfonic acid
MRC	Medical Research Council
mRNA	Messenger ribonucleic acid
MS	Mass spectrometry
MS/MS	Tandem mass spectrometry
MS_M	Mean sums of squares
MS_R	Residual mean sums of squares
mV	Millivolts
MW	Molecular weight
N	Number
N	Non-deleted allele (*CFHR3*/*CFHR1* deletion)
NCBI	National Center for Biotechnology Information
NED	Normal equivalent deviates
NFT	Neurofibrillary tangles
ng	Nanogram
nl	Nanolitre
nLC	Nano-liquid chromatography
nm	Nanometre
O.R.	Odds ratio
^oC	Degree Celsius
OMIM	Online Mendelian inheritance in man
P	Probability
PAGE	Polyacrylamide gel electrophoresis
PBS	Phosphate buffered saline
PCR	Polymerase chain reaction
pg	Picogram
pH	Potential of hydrogen (*potentia hydrogenii*)
pM	Pico molar
psi	Pound per square inch (unit of pressure)
PVDF	Polyvinylidene fluoride
qPCR	Quantitative PCR
QT	Quantitative trait
QTL	Quantitative trait loci
R	Arginine (amino acid)
R^2	Squared correlation

RFLP	Restriction fragment length polymorphism
RHO	Rhodopsin
RNA	Ribonucleic acid
RPE	Retinal pigment epithelium
rpm	Revolutions per minute
rs#	Reference SNP
SCR	Short consensus repeats
SD	Standard deviation
SDS	Sodium dodecyl sulfate
SDS®	Sequence Detection System
SE	Standard error
SELDI	Surface enhanced laser desorption/ionisation
SILAC	Stable isotope labelling by amino acids in cell culture
SNP	Single nucleotide polymorphism
SOLAR	Sequential Oligogenic Linkage Analysis Routines
SPSS	Statistical Package for the Social Sciences
SS_M	Mean sum of squares
SS_R	Residual sum of squares
SS_T	Total sum of squares
STR	Short tandem repeat
T	Thymine (DNA)
T	Tesla (magnetic flux density)
t	Time (minutes)
Taq	*Thermus aquaticus*
TBE	Tris base, boric acid and EDTA
TC	Total cholesterol
TE	Tris, EDTA
TFA	Trifloroacetic acid
TG	Triglyceride
TMB	3,3′,5,5′-tetramethylbenzidine
TOF	Time-of-flight
U.K.	United Kingdom
U.S.	United States
UCSC	University of California Santa Cruz
UF	Ultrafiltration

un	Units
UV	Ultraviolet
v	Version
V	Volts
VIF	Variance inflation factor
vLDL	Very low density lipoproteins
Y	Tyrosine (amino acid)

LIST OF FIGURES

Figure 1.1 Comparison of linkage and association strategies in mapping QTL. 4
Figure 1.2 Modelled distribution of phenotypic effect sizes and the number of underlying disease susceptibility variants. 4
Figure 1.3 Map of the Dalmatian islands of Croatia in the Adriatic Sea. 13
Figure 1.4 Pie chart representing the relative contribution of proteins within plasma. 15
Figure 1.5 The Principle of MALDI-TOF Mass Spectrometry 19
Figure 1.6 The Principle of electrospray ionisation 19
Figure 1.7 The sequence of pathogenic events leading to Alzheimer disease as proposed by the amyloid cascade hypothesis. 27
Figure 1.8 Proteolytic processing of APP. 29
Figure 1.9 Risk of Alzheimer disease by quartiles of plasma Aβ42 concentration. 32
Figure 1.10 Hazard ratios for dementia by concentrations of Aβ40 and Aβ42. 32
Figure 1.11 The hypothetical isoform-specific effects of APOE on neuronal repair, remodelling and protection. 40
Figure 1.12 The structure of the human eye. 42
Figure 1.13 Regulation of the Cleavage of C3 by Factor H and Factor I. 46
Figure 1.14 Results of the AMD association analysis of *CFH* SNPs carried out by Li *et al*., 2006. 49
Figure 1.15 Project outline 55
Figure 2.1 Illustration of the *APOE* genotyping. 64
Figure 2.2 Diagrammatic presentation of the quantitative PCR methodology. 71
Figure 2.3 BLAST2 sequence analysis of the *CFHR4* and *CFHR3* genes. 73
Figure 2.4 Optimisation results of probe, primer and DNA template concentrations. 73
Figure 2.5 Transformation of cumulative percentages into normal equivalent deviates. 96
Figure 3.1 Predicted power for detecting polymorphisms of small to large effect size, influencing a quantitative trait by genetic association. 117
Figure 3.2 Analysis of power in a case-control genome-wide association study. 125
Figure 4.1 Diagrammatic representation of the “sandwich” ELISAs of plasma Aβ40 and Aβ42. 130
Figure 4.2 Standard curves of the two commercial assays of Aβ40 and Aβ42 in plasma. .. 131
Figure 4.3 Plasma Aβ measurements in general population samples from the island of Vis. 134
Figure 4.4 Results of plasma Aβ40 genome-wide linkage scan. 139
Figure 4.5 Results of plasma Aβ42 genome-wide linkage scan. 140
Figure 4.6 Results of genome-wide linkage scan for the ratio of Aβ42:Aβ40 in plasma. 141

Figure 5.1 Results of the pedigree-based genome-wide quantitative trait locus association analysis of plasma Aβ40 in Vis. 151
Figure 5.2 Analysis of the chromosomal region 9q33 showing suggestive evidence of association with plasma Aβ40. 154
Figure 5.3 Analysis of plasma Aβ40 association and linkage disequilibrium between SNPs at the 9q33 locus. 155
Figure 5.4 Results of the genome-wide quantitative trait locus association analysis of plasma Aβ42 in the Vis population. 158
Figure 5.5 Analysis of plasma Aβ42 association and linkage disequilibrium between SNPs at the 2q32 locus. 161
Figure 5.6 Results of the genome-wide quantitative trait locus association analysis of plasma Aβ42:Aβ40 ratio in the Vis population. 164
Figure 5.7 Analysis of the suggestive association peak at 2q22 in relation to the Aβ42:Aβ40 ratio in plasma. 167
Figure 6.1 Diagrammatic presentation of “sandwich” ELISA of CFH. 178
Figure 6.2 Testing the purity of the CFH standard and the specificity of the OX23 antibody. 178
Figure 6.3 Titration curves showing the stages of CFH assay optimisation. 181
Figure 6.4 Effect of varying incubation times with TMB chromogen. 182
Figure 6.5 Analysis of inter and intra-experimental variation of CFH ELISA assay 183
Figure 6.6 Analysis of variation after periods of sample freeze-thawing 183
Figure 6.7 Effect of different pipetting techniques on assay variation 185
Figure 6.8 Plasma CFH measurements in four study populations. 187
Figure 6.9 Plasma CFH covariate analysis in the isolate Dutch population. 191
Figure 6.10 Analysis of plasma CFH covariates in an urban Scottish population. 193
Figure 6.11 Analysis of plasma CFH concentrations between Scottish AMD cases and controls. 193
Figure 6.12 Results of plasma CFH linkage analysis in the Vis population. 198
Figure 7.1 Diagrammatic representation of the *CFH* locus. 206
Figure 7.2 Genotyping of the *CFHR3/CFHR1* deletion by quantitative real-time PCR. 208
Figure 7.3 Screening for individuals homozygous for the *CFHR3/CFHR1* deletion. 208
Figure 7.4 Results of the genome-wide quantitative trait locus association analysis of plasma CFH in the Vis population. 212
Figure 7.5 High resolution view of the Regulator of Complement Activation (RCA) cluster on 1q31 and SNPs associated with plasma CFH concentration in Vis. 214
Figure 7.6 Differences in plasma CFH concentration associated with rs6677604 genotypes. 214

Figure 7.7 Analysis of linkage disequilibrium across the *CFH* locus based on the analysis of 492 unrelated individuals from Vis island..... 218
Figure 7.8 Plasma CFH association test and linkage disequilibrium analysis in the Dutch (Rucphen) isolate. 226
Figure 7.9 Test of association in unrelated cases of AMD and controls. 230
Figure 7.10 The relationship between AMD odds ratios and the effects on plasma CFH concentration associated with single variants. 232
Figure 7.11 Logistic regression analysis of AMD susceptibility factors in the Scottish case-control cohort. 233
Figure 7.12 Sequence analysis of the 354.5 kb region genetically associated with plasma CFH concentration. 241
Figure 8.1 Mass spectra of crude and immunodepleted human plasma. 255
Figure 8.2 MALDI-TOF MS spectra of the small mass range showing results of ZipTip treatment of plasma. 258
Figure 8.3 MALDI-TOF MS spectrum of human plasma after ultrafiltration using 10K MW cut-off columns. 259
Figure 8.4 Assessment of MALDI-TOF MS in quantitative determination of Aβ42 peptide. 261
Figure 8.5 Assessment of the MALDI-TOF technique in quantitative determination of differences in Aβ42 concentrations after ZipTip treatment. 263
Figure 8.6 MALDI-TOF MS results of the human plasma medium size mass range after immunodepletion and hydrophobic interaction chromatography. 264
Figure 8.7 MALDI-TOF MS results of the large mass range after immunodepletion and ZipTip treatment of crude human plasma. 266
Figure 8.8 Results of capillary electrophoresis of varying concentrations of tryptophan..... 268
Figure 8.9 Capillary electrophoresis results of varying the length of injection time on the observed tryptophan peak area. 269
Figure 8.10 Plasma protein staining prior to ion exchange chromatography. 273
Figure 8.11 Comparison of CFH concentrations measured by ELISA and IEC. 275
Figure 8.12 Test of correlation between ELISA and IEC measurements of plasma CFH... 276
Figure 9.1 A simplified diagrammatic presentation of a eukaryotic gene during transcription. 296
Figure 9.2 A proposed model for the relationship between AMD susceptibility factors including those affecting plasma CFH concentration based on the findings of the present study. 298

LIST OF TABLES

Table 1.1 Common techniques for the assay and detection of plasma proteins 16
Table 1.2 Summary of the techniques used in identification and characterisation of proteins in complex mixtures. 21
Table 1.3 Transgenic rodent models of Alzheimer disease pathology 35
Table 1.4 A summary of conclusions from meta-analysis of established associations between AMD and genetic variants. 53
Table 2.1 Descriptive statistics of sample datasets. 61
Table 2.2 Composition of the PCR master mix used in the amplification of *APOE* DNA fragment. 63
Table 2.3 Composition of the CfoI digest for *APOE* genotyping..... 65
Table 2.4 Expected fragment sizes of amplified *APOE* after digestion with *Cfo*I. 65
Table 2.5 Composition of the BigDye sequencing reaction mix. 66
Table 2.7 Chemical composition of the Laemmli gels used in one-dimensional protein separation. 82
Table 3.1 *APOE* Allele and genotype frequencies in Vis. 112
Table 3.2 Univariate analysis of *APOE* genotypes and plasma lipids concentrations. 115
Table 3.3 Comparison between reported estimates of heritability of plasma lipid concentrations and estimates obtained in the Vis population. 120
Table 3.4 Comparison of *APOE* association with plasma lipid concentrations between Vis and published data. 120
Tables 4.1-4.3 Analysis of covariate effects associated with concentrations of Aβ40, Aβ42 and Aβ42:Aβ40 ratio in plasma. 136
Table 4.4 Summary of results from the genome-wide linkage analysis in the Vis population 142
Table 5.1 Summary of the results of a genome-wide association analysis of plasma Aβ40 in the Vis population..... 153
Table 5.2 Summary of the results of whole-genome association analysis of plasma Aβ42 in the Vis population..... 159
Table 5.3 Summary of the genome-wide QTL association analysis of the plasma Aβ42:Aβ40 ratio in the Vis population. 165
Table 5.4 SNP coverage of the genes involved in familial Alzheimer disease by the Illumina Hap300 array. 174
Table 6.1 Assay variation observed between different time-points. 183
Table 6.2 Plasma CFH covariate analysis in Vis. 188
Table 6.3 Descriptive statistics for rs1061170 (*CFH* Y402H) in the Vis population..... 188
Table 6.4 Descriptive statistics for SNP rs1061170 in the Dutch Rucphen population. 191

Table 6.5 Analysis of plasma CFH concentration in the Japanese AMD cohort. 195
Table 7.1 Genotype and allele frequencies for the deletion of the *CFHR3/CFHR1* in four population sample sets. 210
Table 7.2 Details of additional (non-Illumina) *CFH* SNPs genotyped in the Vis (Croatian) sample set. 210
Table 7.3 Summary table of SNPs showing evidence of genome-wide association with plasma CFH. 212
Table 7.4 Summary of the polymorphisms showing full and suggestive associations with plasma CFH after correction for multiple testing. 216
Table 7.5 Results of the linear regression test for collinearity of markers used in the analysis of plasma CFH. 221
Table 7.6 Results of the univariate analysis of plasma CFH using a multi-marker model obtained by 'backward' regression analysis 221
Table 7.7 Association analysis of the 3-SNP haplotypes with plasma CFH concentration in the Vis population. 223
Table 7.8 Single marker associations with plasma CFH concentration in the Dutch (Rucphen) isolate. 227
Table 7.9 AMD case-control association analysis in the Scottish cohort. 229
Table 7.10 Association between plasma CFH and AMD in the Scottish cohort. 232
Table 7.11 Association analysis of 4-SNP haplotypes from Li *et al*. (2006) in Scottish AMD case-control samples. 236
Table 7.12 AMD case-control association analysis of unrelated Japanese individuals. 238
Table 7.13 Overview of the data published on genetic association analysis of the *CFH* locus. 246
Table 7.14 Comparison of the minor allele frequencies between the Scottish and Japanese control populations. 250

CHAPTER 1

INTRODUCTION

1.1 Mapping genes for common, complex human diseases

The recent completion of the human genome sequence and rapid improvements in genotyping technologies can allow the genome-wide analysis of genetic variation influencing complex disorders (Mayeux, 2005). Some of these variants may have a direct effect on disease, and some may influence quantitative traits which are risk factors for disease. Success has been achieved in mapping genes for Mendelian diseases (Thomas & Kejariwal, 2004). However, unravelling the complex nature of common diseases has proven more challenging. The main differences between Mendelian and complex disorders and some of the techniques employed to study them are summarised below.

Mendelian and complex disorders

In Mendelian disorders, disease is typically caused by one of several genes, each with many rare alleles (Wright *et al*., 1999). These are usually highly penetrant deleterious alleles with high allelic heterogeneity and low disease prevalence which segregate in families. In the case of more genetically complex disorders, a combination of intermediate phenotypes, each under genetic control, affects predisposition to disease. Common, complex diseases characterised by late age of onset are a manifestation of many genetic variants and loci with incomplete penetrance which interact with one another (epistasis) and with the environment. It is often preferable to study the intermediate disease endpoints rather than categorical diseases as the number of susceptibility loci may be so high and the disease state can be a complex and insensitive indicator of underlying pathogenic processes (Wright *et al*., 1999)

A number of factors have resulted in shifts in the selection pressure acting on the biological processes that underlie major classes of common disease (Di Rienzo, 2006). Crucial factors include the dispersal of modern humans out of Africa, and a shift away from a hunting and gathering life-style in addition to the more recent transitions associated with industrialisation and globalisation. These have resulted in major changes in environmental factors which are believed to influence the evolution of susceptibility variants for common diseases (Di Rienzo, 2006).

Linkage and LD mapping strategies

Linkage and association mapping are the two main genome-wide approaches for mapping genes that underlie common diseases and their associated quantitative traits (QTs).

Genome-wide linkage analysis is the traditional method of identifying disease genes and has proven successful for mapping genes that underlie monogenic disease. The principle is based on the segregation of genetic markers flanking the disease gene with the disease in families (Blangero, 2004). Genome-wide linkage analysis has also been used in mapping quantitative trait loci (QTL) with the aim of finding chromosomal regions with evidence of QTL-specific heritability and inferring the existence of a QTL. However, limited success has been achieved for most complex traits using linkage and the few genes discovered in this way only explain a small proportion of the overall trait heritability (Altmuller *et al*., 2001). This lack of power can be due to the relatively low heritability of complex traits and the inability of the standard set of microsatellite markers (5-10cM apart) to extract complete information about inheritance (Hirschhorn & Daly, 2005). The complex architecture of common diseases (described in the next section) and the presence of common genetic variants of small to modest effect also make linkage studies less powerful.

Genome-wide association studies have become possible since the advent of high density single nucleotide polymorphism (SNP) genotyping platforms. A SNP is a single nucleotide substitution with a minor allele frequency of 0.01 or greater. In quantitative trait association analyses, the effect of a marker genotype on the distribution of the QT is directly assessed using a linear regression model (Blangero, 2004). Success in this type of study will depend on the magnitude of linkage disequilibrium (LD) between the genetic marker and the causal variant in the population. Linkage disequilibrium is defined as the non-random association of alleles at two or more loci, not necessarily on the same chromosome (Cannon, 1963). The LD relationship is highly dependent on the allele frequencies of the marker and of the functional variant, a feature that cannot be predicted (Blangero, 2004). Larger sample sizes can compensate for the negative effect of low LD and mismatch in QTL allele frequencies (**Figure 1.1**).

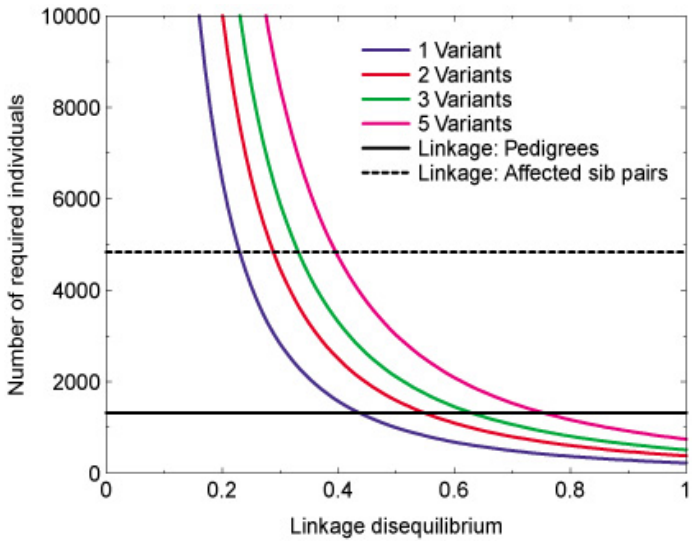


Figure 1.1 Comparison of linkage and association strategies in mapping QTL.

Number of individuals required to achieve 80% power to detect a QTL. The Model is based on the assumption that the QTL explains 15% of the trait variance and the prevalence of the disease is 25%. The association approach is highly dependent on the strength of LD between genetic marker and QTL and is shown to be more powerful than the linkage approach for both pedigrees (when LD > 0.20) and affected sib-pairs (when LD > 0.40). The association approach is successful if detecting up to five quantitative trait nucleotides (QTN) with equal relative effects, when a moderately large sample size is used. However, linkage is more powerful in many circumstances when such as high penetrance null alleles even in the presence of rather high LD. Figure adapted from Blangero, 2004.

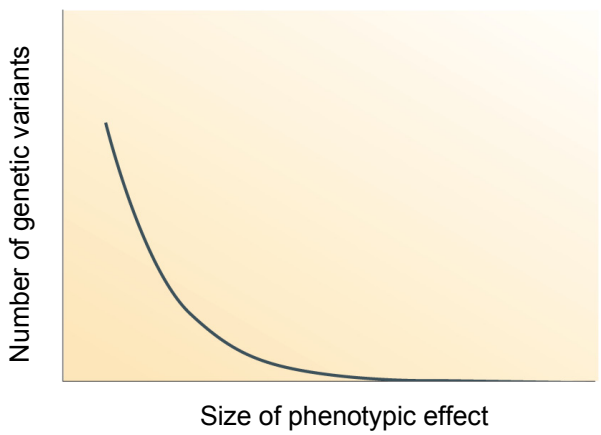


Figure 1.2 Modelled distribution of phenotypic effect sizes and the number of underlying disease susceptibility variants.

A model for the exponential or L-shaped distribution of the allelic effect sizes associated with complex disease susceptibility variants. It suggests that the underlying genetic variation, which affects complex traits, is based on a small number of variants with large effects and a large number of variants with small effects. Figure adapted from Wang *et al.*, 2005.

When comparing two contrasting linkage designs (sib-pairs and large pedigrees) the extended pedigree design is up to four times more efficient than the sib-pair design as has been shown in family-based linkage studies of thrombosis (Blangero *et al*., 2003). Assuming there is a single variant accounting for 15% of the trait's variance, the association approach using unrelated individuals is more powerful than the affected sib-pair design for all LD correlations greater than 0.20 (Blangero, 2004). However, for LD correlations smaller than 0.20 the affected sib-pair design would have more power in detecting a single underlying polymorphism. For LD correlations smaller than 0.4 the extended pedigree approach would be more efficient in detecting a single variant and similarly for LD correlations smaller than 0.75, this approach would have enough power to detect five variants associated with the trait. Therefore, linkage is more powerful in many circumstances such as the presence of highly penetrant null alleles even with high LD correlations. Although association studies are prone to population stratification, they represent an unbiased approach that can be attempted even in the absence of evidence regarding the function or location of the causal genes (Blangero, 2004).

1.1.1 Genetic architecture of complex traits

An important route for complex disease mapping is through the study of normal physiological variation of quantitative trait risk factors (Falchi *et al*., 2004). Intermediate phenotypes, which are assumed to be functionally related to a broader phenotype such as disease, may have stronger genetic determinants than the downstream phenotype they mediate, contributing to the added power of QT analysis over binary traits (Falchi *et al*., 2004).

Quantitative genetics theory assumes an explicit genetic model to describe the genetic architecture of a complex trait. A comprehensive understanding of the genetic architecture of any complex trait requires knowledge of several factors; the numbers and identities of the genes which lead to the trait phenotype; their effect sizes and allele frequencies; the mutation rates at these loci; all two-way and higher order epistatic interaction effects; the presence of pleiotropic effects on other traits, especially reproductive fitness; finally, the molecular mechanisms causing the differences in trait phenotype (MacKay, 2001).

Two main models have been proposed to account for the allelic architecture of common diseases (Wang *et al*., 2005). First, the common disease/common variant hypothesis (CDCV) suggests that susceptibility to common diseases is a result of the joint action of several common variants, so that unrelated affected individuals share a significant proportion of disease alleles in common (Reich & Lander, 2001). The alternative is the classical disease heterogeneity model, which suggests that disease susceptibility is mainly due to distinct genetic variants of low minor frequency (MAF < 0.01) that are present in different affected individuals (Smith & Lusis, 2002).

The distribution of genetic effect sizes conferred by individual variants also has an effect on the allelic architecture of complex traits. Studies of *Drosophila melanogaster* and animal models have indicated that the distribution of phenotypic effect sizes of genetic variants is consistent with the existence of a small number of genetic loci with large effects and numerous loci with small effects (**Figure 1.2**).

1.1.2 Lessons from association/linkage mapping of QTL

The lack of success of linkage analyses can either be due to the small size of the study population (i.e. size of the pedigrees) or the lack of power in linkage to detect variants of modest effect size as reflected by a low QTL-specific heritability (Dawn & Barrett, 2005). The power in genome-wide analyses has been shown to be directly related to the QTL heritability in the case of association but to the square of the QTL heritability in the case of linkage (Sham *et al*., 2000). Therefore, it is expected that with decreasing QTL heritability, association will become progressively more powerful than linkage. Using simulations, Sham and colleagues (2000) have estimated that approximately 20,000 sib-pairs would be required to detect a QTL of 10% effect size (i.e. explaining 10% of the QT variance) using linkage with 80% power. The number of sib-pairs decreases to approximately 1,000 for detection of statistically significant associations, assuming that the degree of linkage disequilibrium between the causal QTL and the marker is strong ($R^2 > 0.50$).

For linkage, the level of significance required is set at a LOD score of 3 for monogenic traits, which is equivalent to a fixed-sample one-tailed significance level of 0.0001 (Morton, 1955). The LOD score is defined as the common logarithm of the likelihood ratio (1,000) that is necessary to convert the odds from 50:1 against

linkage to 20:1 in favour of linkage (Ott, 1991). For complex traits, slightly higher LOD score thresholds are required (Lander & Kruglyak, 1995). Interpretation of significance levels in genome-wide association studies aims to keep the false positive rate resulting from small sample sizes and population stratification within acceptable bounds and leads to the view that very low *P*-values are needed for strong evidence of association (Wellcome Trust Case Control Consortium (WTCCC), 2007). For a given significance threshold, the probability of a true association depends on the prior odds and the power (Wacholder *et al*., 2004). Interpreting the strength of evidence in an association study depends on the likely number of true associations and the power to detect them, which in turn depends on effect sizes and sample size (WTCCC, 2007).

Hidden population structure is a cause of false-positive findings due to confounding when different ancestries carry higher disease risk and are, as a result, over-represented in cases or between subgroups in the population (WTCCC, 2007). For example, a genome-wide study of seven common diseases has shown that the British population is heterogeneous, identifying thirteen genomic regions showing strong geographical variation in allele frequencies (WTCCC, 2007).

The statistical power to detect causal variants increases with increasing allele frequency and penetrance, as reflected by the proportion of the phenotypic variance explained by the variant (Hirschorn & Daly, 2005). Mapping by association requires low genetic and environmental heterogeneity (Kruglyak, 1997). Although linkage approaches are less affected by allelic heterogeneity, they lack power to detect variants with small effect (Risch & Merikangas, 1996; Camp, 1997). The most efficient method may be to search for linkage and association jointly (Terwilliger & Weiss, 1998). If allelic association is present, it can increase the power of a linkage test, if it is not present, linkage remains a viable approach for low-resolution mapping (Wright *et al*., 1999).

1.1.3 Mendelian randomisation

A setback in genetic epidemiological analyses is the lack of consistency between observational studies and randomised controlled trials. This arises when potentially causal factors are associated with a disease-related trait but this is not confirmed

when tested in randomised controlled trials. The most likely explanation is the presence of a confounding factor (Hill, 1965). By definition, a confounding variable is associated with both the probable cause and the outcome. The principle of Mendelian randomisation is to make use of genetic markers that alter the level of a modifiable environmental exposure that is associated with disease risk. Genetic variants of this sort should be associated with disease risk if the association is causal but not if it is due to confounding (Davey *et al*., 2005). Therefore, genetic polymorphisms influencing a disease risk factor can be used to study the causal relationship between an intermediate and disease risk. The use of genetic markers in this way has several advantages (Davey *et al*., 2005). Firstly, genetic variants are not generally affected by behavioural and physiological factors. Furthermore, such variants will not be associated with other variants, except as a result of linkage disequilibrium (LD). Secondly, genetic variants are not influenced by reverse causality i.e. they never occur as a result of the phenotype they control. For example intermediate phenotypes such as plasma cholesterol, CRP and fibrinogen can be associated with coronary artery disease but it is not clear whether this is due to a causal effect. Genetic markers can be advantageous over biological markers which may be influenced by disease status (reverse causality). Thirdly, genetic markers are less prone to measurement error and within-individual variability compared with intermediate phenotypes.

The major limitation of Mendelian randomisation is the difficulty in establishing a reliable association between genetic markers and either disease-related intermediate phenotypes or disease itself (Davey & Ebrahim, 2003). This can be difficult because of population stratification, small sample size and lack of suitable polymorphisms for studying modifiable exposures of interest.

1.1.4 Choice of population

Association analyses are dependent on the extent of LD between the marker and QTL of interest. The number of generations since the common ancestor or the number of meioses during this time are the main causes of the decay in linkage disequilibrium, and are related to the number of recombination events and the effective population size. Recombination equilibrates linked alleles, whereas small

population size results in the generation of new allelic disequilibria by genetic drift. Genetic drift is defined as the random sampling of mutations from generation to generation resulting in chance changes in allele frequencies (Hartl & Clark, 1997).

Different types of population have been successfully used in mapping genes influencing human diseases. Recent founder populations such as Finland, Iceland, Sardinia and Japan have been useful in finding rare Mendelian genes (Laan & Paabo, 1997). In these populations allelic heterogeneity may have diminished due to genetic drift before population expansion. However, due to the small allelic effect sizes, they have not proven as useful in mapping complex disease genes (Wright *et al*., 1999). Small populations that have remained stable in size such as the Scandinavian Saami may prove to be more successful in LD mapping due to the presence of large extents of LD caused by genetic drift as shown by experimental data on the X-chromosome (Laan & Paabo, 1997).

Isolated inbred populations such as those found in Switzerland (Ellis & Starmer, 1978), Sardinia (Workman *et al*., 1975) and North America (Bear, 1988) also have potential in LD mapping. Inbreeding has been shown, in a simulation study, to increase the power to detect susceptibility alleles in complex disorders (Genin & Clerget-Darpoux, 1996). Furthermore, genome-wide linkage mapping has successfully mapped loci influencing plasma triglyceride concentration in a single 1,623-member pedigree from the Hutterite population (LOD > 3, $P < 0.05$) despite lack of success in other populations (Newman *et al*., 2003). This is due to reduced effective population size, resulting in reduced genetic heterogeneity and more extended regions of LD as a result of recent founding effects or genetic drift.

Genetically simplified isolated populations provide optimal conditions for LD mapping (Wright *et al*., 1999). Isolation, due to geographical or cultural factors, results in a more uniform distribution of environmental risks as well as reduced genetic heterogeneity. Furthermore, increased inbreeding in these populations can influence both disease prevalence and its underlying quantitative traits.

1.1.4.1 Gene mapping in isolated populations

A powerful approach to mapping genes influencing complex traits is to study isolated founder populations. Genetic heterogeneity is likely to be reduced,

depending on the original composition of the particular isolate and the number of founders. Environmental variation is also likely to be reduced because of uniform social behaviour, for example absence of cigarette smoking in the Hutterite population for religious reasons (Newman *et al*., 2003). In isolates, extended genealogical data can often be readily obtained facilitating the identification of identity-by-descent relationships in mapping. A population may be referred to as an "isolate" if it has not had a large degree of admixture with surrounding populations for a given number of generations due to either geographic or socio-political barriers (Escamilla, 2001). If an isolate does not admix with its neighbours there is a chance that particular founder mutations may be present in a large number of members of the current generation. This may be the result of normal growth of the population, selection or random genetic drift.

In a simulation study of 1,053 related individuals in 4-5 generation pedigrees from a Dutch isolate, large fluctuations were observed in allele frequencies when the initial allele frequencies were lower than or equal to 1%; and smaller fluctuations when the initial frequencies were larger than 1% (Pardo *et al*., 2005). Comparing empirical data from the Dutch and Icelandic isolates and the CEPH outbred population, Pardo and colleagues showed that although large fluctuations in allele frequencies might occur in recent isolates due to drift and founder effects, common genetic variants with a frequency higher than 1% are expected to be present in both young genetic isolates and outbred populations.

Successes in mapping susceptibility loci for complex disorders have been achieved in some European isolate populations. In the Icelandic population, common susceptibility variants have been mapped for prostate cancer (Gudmundsson *et al*., 2008; Gudmundsson *et al*., 2007), abdominal aortic aneurysm and intracranial aneurysm (Helgadottir *et al*., 2008), glaucoma (Thorleifsson *et al*., 2007), periodic limb movements in sleep (Stefansson *et al*., 2007), atrial fibrillation (Gudbjartsson *et al*., 2007), breast cancer (Stacey *et al*., 2007; Stacey *et al*., 2006), myocardial infarction (Helgadottir *et al*., 2007) using genome-wide association. QTL for migraine (Bjornsson *et al*., 2003), osteoarthritis (Stefansson *et al*., 2003), Parkinson disease (Hicks *et al.*, 2002), stroke (Gretarsdottir *et al*., 2002) and essential tremor (Gulcher *et al*., 1997) were mapped using genome-wide linkage. All the above

studies have involved large numbers of related individuals and the results have been replicated in Icelandic, U.S. or other European populations. Furthermore, the effect sizes observed range between odds ratios of 1.1 to 4.9 (moderate to large effect size). However, the success in mapping common variants for common traits in this population could be purely due to the large number of samples available and well-designed studies. It has been proposed that the Icelandic population is not a homogeneous genetic isolate, but instead was founded by a large mix of populations of Norse and Gaelic origin (Arnason *et al*., 2000). Therefore, it is important that investigators provide evidence that their samples are from truly isolated populations (Heutink & Oostra, 2002).

A common susceptibility locus for asthma-related traits has been mapped to chromosome 7p using linkage followed by fine-scale association mapping in a series of Finnish patients and controls from Kainuu (Laitinen *et al*., 2004). The assumption was made that founder effects in this population might aid the identification of common variants influencing such common traits Laitinen and colleagues identified a G-protein-coupled receptor for asthma susceptibility which lies within a 133-kb genetic segment containing an asthma-predisposition haplotype. The findings of the study were simultaneously replicated in two other populations with histories of a founder effect, Quebec (Canada) and North Karelia (Finland).

Fine-mapping by association of a previously-identified region of linkage on chromosome 1q21-1q23 (Pajukanta *et al*., 1998) in 60 extended Finnish families with familial combined hyperlipidemia from a relatively isolated population, identified the susceptibility locus Upstream Transcription Factor 1 (*USF1*), which is observed in about 20% of individuals with premature coronary heart disease (Pajukanta *et al*., 2004). The effect size of the associated variant was found to be significantly larger in males with higher plasma triglyceride concentration. However, the findings of this study were not replicated in another population and it was not clear whether using an isolated population facilitated the identification of the associated locus.

A study consisting of 44 families (775 individuals) from the isolated Sardinian village of Talana, identified a 15 cM (~15 Mb) region on chromosome 2q21-22, involved in determining total plasma cholesterol and low-density lipoprotein cholesterol, using genome-wide variance-component linkage analysis

(Falchi *et al*., 2004). The magnitude of effect ranged from LOD score of 3.9 (LDL cholesterol) to 4.3 (total cholesterol). The findings of the study were simultaneously replicated in an isolated population from the same region of Sardinia, the village of Perdasdefogu. The results of the study by Falchi and colleagues replicated the findings of the genome-wide linkage scan, conducted in another isolated population (Hutterite). This study involved 485 individuals in a 1,623-member pedigree enriched for cardiovascular disease, in which a locus with a multipoint LOD score of 3.40 was identified for TG in a nearby linked region (Newman *et al*. 2003).

Although isolated populations can benefit gene localisation, findings made in isolated populations might not be valid in the general population, especially in very old isolated populations where new mutations may arise or old mutations remain while becoming extinct in other populations. Younger isolates may not suffer from this drawback, since they are still genetically similar to the general population because of their recent separation (Heutink & Oostra, 2002).

1.1.5 The Croatia project

The principal aim of the Croatia Project is to evaluate the use of isolate populations for the identification of genes which affect quantitative traits underlying the predisposition to common diseases. This is based on the assumption that the choice of study population and use of quantitative traits rather than categorical diseases are both critical factors in mapping susceptibility genes influencing complex diseases (Wright *et al*., 1999). Croatia has 15 islands in the Adriatic Sea with populations greater than 1,000 (**Figure 1.3**). The villages on the islands have unique histories and have preserved their isolation from other villages and the outside world through many centuries. The history, demography and genetic structure of these villages have been investigated for more than fifty years mainly by the Institute for Anthropological Research in Zagreb (Rudan *et al*., 1987).

In 2002, ten village settlements were carefully selected by Dr. I. Rudan and colleagues to present a wide range of differing ethnic histories, past admixture and bottleneck events and known founding times, in order to examine the unique characteristics of multiple isolated Dalmatian island populations.

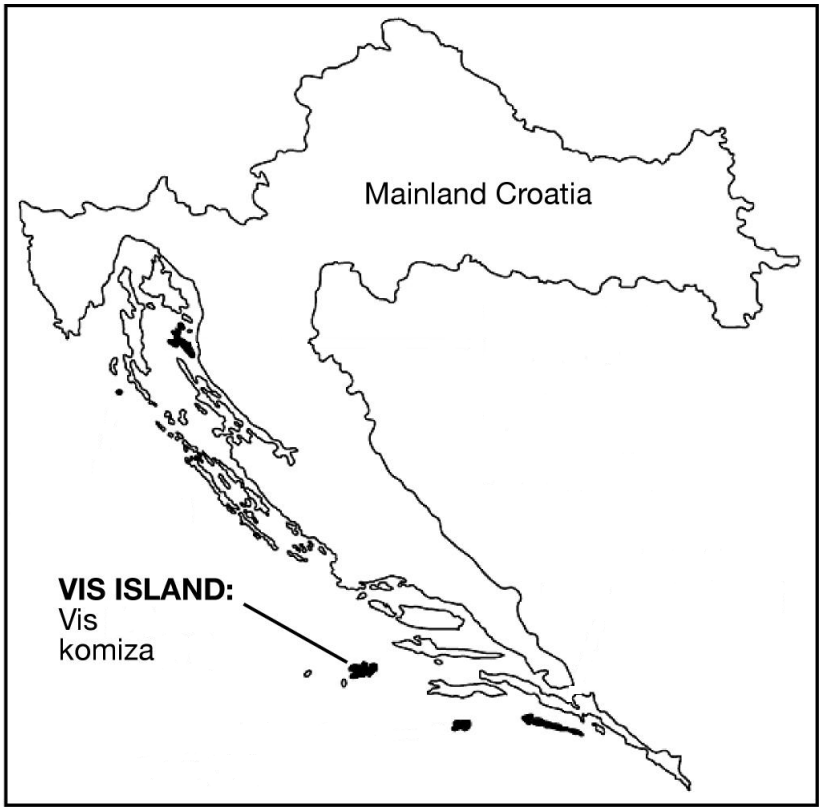


Figure 1.3 Map of the Dalmatian islands of Croatia in the Adriatic Sea.

Geographic location of the island of Vis, from which samples used in this study were obtained. Other islands of the Eastern Adriatic, Northern and Middle Dalmatia, and mainland Croatia are also shown. Figure adapted from Vitart *et al*., 2006.

A preliminary field survey of 10 Croatian isolates was then carried out in 100 individuals from each isolate to clarify their suitability for the genetic analysis of complex traits (Vitart *et al*., 2006). A high level of differentiation and structure is observed between most of the island communities, which was likely to be the result of strong isolation and endogamy (Vitart *et al*., 2006). Furthermore, the extent of LD compared with the urban UK population correlated significantly with the extent of differentiation and was also very high in most of the populations. The extent of LD was assessed using eight linked markers on Xq13-21, a region of low recombination (0.25 cM/Mb) which has been has been used to explore population-specific differences in LD. Markers in that chromosomal region consistently displayed increased pair-wise association in populations with a history compatible with a reduced effective population size (Kaessmann *et al*., 2002; Vitart *et al*., 2005).

To analyse a diverse range of quantitative traits, blood samples were obtained from 1,042 adult volunteers from the villages of Komiza (N = 584) and Vis (N = 458) in 2003, both located on the island of Vis (**Figure 1.3**). Clinical histories and questionnaires were administered and anthropometric and physiological measures were obtained for a range of more than fifty QTs, including blood pressure, lung function, clotting factors, lipids and other clinical biochemistry measures on fasting blood. The availability of plasma samples provided a rich resource for the analysis of biochemical quantitative traits with potential roles in disease susceptibility. Successful studies have been conducted involving the analysis of differentiation and structure in the isolated islands of Dalmatia (Vitart *et al*., 2006), analysis of personality and psychological traits (Ivkovic *et al*., 2007) and the identification of SLC2A9 as a novel urate transporter using a genome-wide association approach (Vitart *et al*., 2008).

1.2 Proteomics and the human plasma

Human blood plasma is generally the most informative proteome from a medical viewpoint since it is the primary clinical specimen and represents one of the largest and most clinically accessible collections of proteins in bodily fluids. An estimated ten thousand different proteins typically exist in the human plasma, twenty-two of which constitute about 99% of the total protein content of plasma while the remaining fraction are low abundance proteins and peptides (Adkins *et al*., 2002) (**Figure 1.4**). Almost all body cells communicate with the plasma either directly or through tissues/biological fluids, and many of these cells release at least a part of their contents into the plasma upon damage or death (Fujii *et al.*, 2005). However, characterisation of the human plasma proteome may pose one of the greatest challenges. Plasma contains low abundance proteins, mainly products of tissue leakage at pg/ml levels, in the presence of several relatively dominant, high abundance proteins such as serum albumin at 35-50 mg/ml (**Figure 1.4**).

Human serum albumin is a single molecule with a molecular weight (MW) of 65 kDa which is the most abundant plasma protein and accounts for approximately 50% of the plasma proteome. It has an important role in the osmo-regularity of the plasma, as well as in transport of many different substances and peptides due to its

specific binding abilities (Campbell & Smith, 1999). Other major components of the highly abundant fraction of plasma proteins include the immunoglobulins (IgG, A, and M, 10-14 kDa, 12-18%), transferrin (75kDa, 2-3%), fibrinogen (2-6%) and α_1-antitrypsin (44 kDa, 2-5%) (Tirumalai *et al*., 2003). Figure 1.4 summarises the relative composition of proteins within the human plasma.

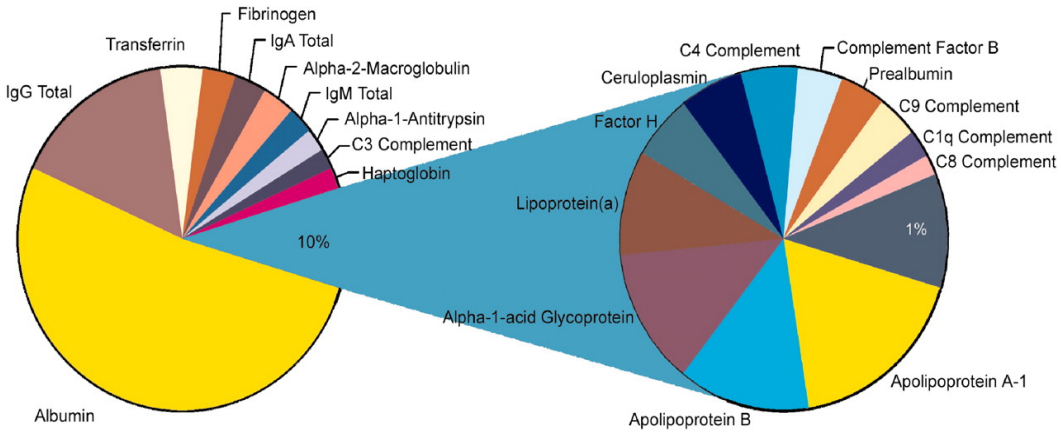


Figure 1.4 Pie chart representing the relative contribution of proteins within plasma.

Twenty-two proteins constitute ~99% of the protein content of plasma, nevertheless there are estimated to be around 10,000 low abundance proteins in plasma (Adkins *et al*., 2002). Diagram adopted from Tirumalai *et al*., 2003.

Human plasma contains a complex array of proteolytically derived peptides (plasma peptidome) that may provide correlates for biological events occurring in the organism (Villanueva *et al*., 2004). Most of these peptides are thought to be fragments of larger proteins that have been partially degraded by endogenous proteolytic enzymes, but their precise identities remain undetermined (Villanueva *et al*., 2004).

1.2.1 Analysis of plasma proteins

The low abundance fraction of the plasma may contain an unexplored archive of useful biomarkers for disease detection. The concentrations of such biomarkers may serve as reliable intermediate phenotypes which are functionally related. However, the identification and analysis of these low abundance proteins and peptides is

hampered by the presence of the more abundant proteins in plasma. To date numerous methods for detection and quantitation of peptides and proteins have been employed as summarised in **Table 1.1**.

Technique	Details
Photometric (absorption)	Commonly used for quantitative assays, simple and rapid. May be affected by colour or turbidity
Photometric (fluorescence)	More sentitive than absorption but more limited choice of fluorescent substrates
Radiometric	Highly sensitive. Problems associated with radioactivity e.g. health hazard and quenching
HPLC	Separation performed rapidly and with high sensitivity. Has not been assessed for quantitation.
Electrochemical	Inexpensive and convenient. Unaffected by colour or turbidity. Not practical for large number of samples.
Electrophoresis	Simple and standard procedure. Semi-quantitative. Limited types of proteins can be assayed, but not a high throughput method.
Immunological	Specific and sensitive for detection and quantitation. High throughput but time-consuming.

Table 1.1 Common techniques for the assay and detection of plasma proteins (Price, 1996).

1.2.1.1 Enzyme-linked immunosorbent assay

Enzyme-linked immunosorbent assay (ELISA) is an immunological detection technique, which offers a highly specific and sensitive assay for detection and quantitation of a wide range of antigens (Engvall & Perlman, 1971; van Weemen & Schuurs, 1971). 'Sandwich' ELISA is the most commonly used form of the assay. In this type of assay, the analyte to be measured is bound between two antibodies: the capture antibody and the detection antibody. The capture antibody is immobilised onto a specially coated surface (e.g. a microtitration plate) and non-specific sites are then blocked. A solution containing the protein or peptide of interest is added and unbound protein is removed by washing and bound antigen is

quantified using a secondary enzyme-linked antibody. Upon addition of substrate, the enzyme produces a coloured product that can be detected photometrically. An important consideration when designing a sandwich ELISA is that the capture and detection antibodies must recognise two non-overlapping epitopes. When the antigen binds to the capture antibody, the epitope recognised by the detection antibody must not be obscured or altered.

The limiting factor in design and use of immunological assays is the availability of antibodies to detect a specific protein. This also implies that ELISAs can only be applied to proteins already identified, and to which antibodies have been raised. Therefore identification of novel proteins is not possible. Moreover, immunological assays would not allow for large-scale quantitation of many plasma proteins as these types of assay are often optimised for detection of one or only a few antigens.

1.2.1.2 Mass spectrometry

The ability to screen and discover multiple biomarkers simultaneously in clinical proteomics has been advanced by the recent success of mass spectrometry (MS). MS consists of two key processes, the generation of gas phase ions of the molecule of interest by ionisation, a process that occurs at the source area of the mass spectrometer, and the analysis of the mass-to-charge ratio (m/z) of these ions using a mass analyser. During the past two decades, MS has become established as the primary method for protein identification from complex mixtures of biological origin, through the analysis of mass to charge ratio of ions.

The main protein ionisation techniques employed in MS proteomic analysis are matrix-assisted laser desorption/ionisation time-of-flight (MALDI-TOF) and electro-spray ionisation (ESI). In MALDI-TOF MS the analyte is dispersed in a solid matrix and spotted onto a so called MALDI target plate (platinum or gold) and ionised through application of a laser beam which is primarily absorbed by the matrix causing desorption and ionisation of the analyte (Karas & Hillenkamp, 1988). The matrix consists of crystallized molecules, of which the three most commonly used are 3,5-dimethoxy-4-hydroxycinnamic acid (sinapinic acid), α-cyano-4-hydroxycinnamic acid (alpha-cyano or alpha-matrix) and 2,5-dihydroxybenzoic acid

(DHB) depending on the type and size of the analyte molecules. Matrix compounds have the following characteristics: firstly, they are of a fairly low molecular weight (to allow facile vaporisation), but are large enough (with a low enough vapour pressure) not to evaporate during sample preparation; secondly, they are acidic, therefore act as a proton source to encourage ionization of the analyte; and thirdly, they have a strong optical absorption in the UV range and rapidly and efficiently absorb the laser irradiation. The ionisation process creates a vaporous mixture of charged particles which is accelerated down a so called flight tube through application of an electric field (**Figure 1.5**). The time-of-flight (TOF) of the ionised particles depends on their mass-to-charge ratio (m/z) and the conversion of the flight times to corresponding m/z is achieved via a quadratic transformation in which the time of flight of the ion varies with the square root of its mass-to-charge ratio (van der Werff *et al*., 2008).

SELDI-TOF MS is a similar technique to MALDI-TOF MS, but the analyte is spotted on a target plate which has been coated to preferentially bind a sub-class of proteins and the unbound molecules are washed off (Hutchens & Yip, 1993). Examples of SELDI-TOF MS application include the analysis of the amyloid-β (Aβ) species in the cerebral cortex using Aβ-specific antibodies (Terai *et al*., 2001) and identification of biomarkers in synovial fluid for rheumatoid arthritis (Uchida *et al*., 2002). During ESI, the liquid sample flows through a thin needle and, by application of high voltage, the small droplets are charged and a very fine spray is generated (Fenn *et al*., 1989). During the passage from the needle to the MS instrument the solvent is evaporated and the density of the charges within the droplets rises, which results in charged analyte molecules (Frohlich & Arnold, 2006) (**Figure 1.6**).

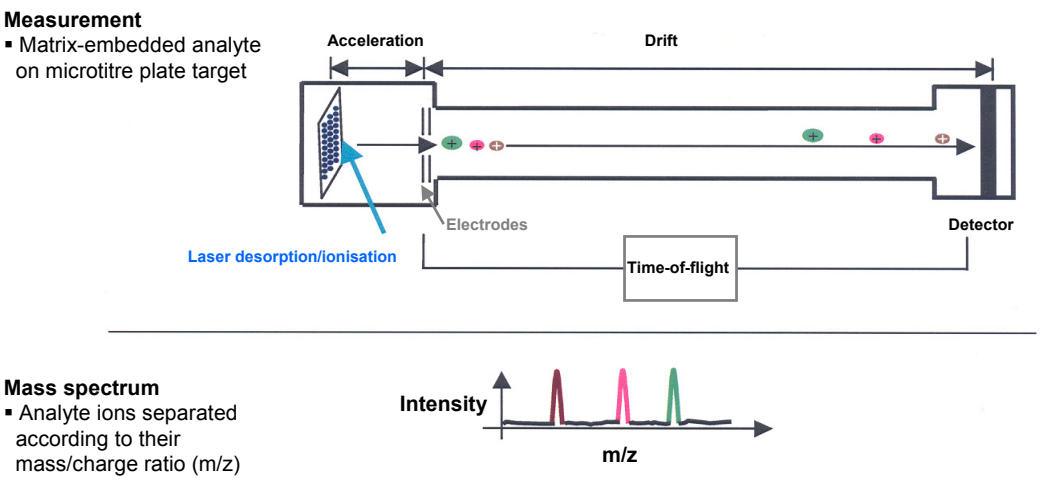


Figure 1.5 The Principle of MALDI-TOF Mass Spectrometry

Bombardment of sample molecules with a laser light brings about sample ionisation. The laser energy is transformed into excitation energy by the matrix compound with which the sample has been mixed. The analyte and matrix ions are then released and the time taken for them to be detected, which is directly proportional to their size, is converted to a mass-to-charge ratio (m/z).

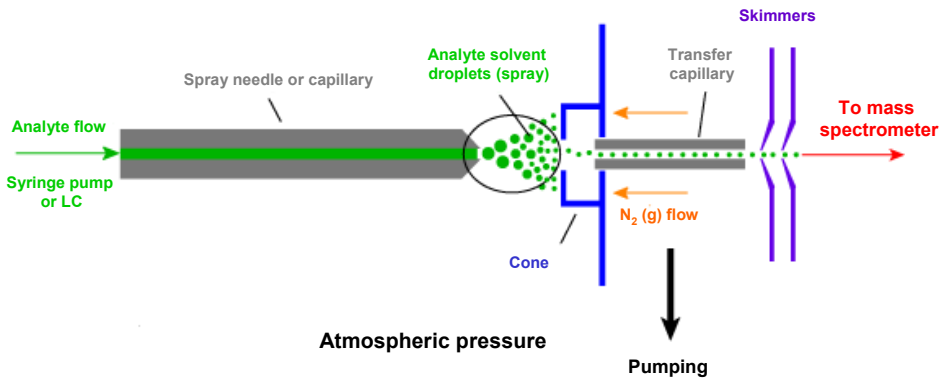


Figure 1.6 The Principle of electrospray ionisation.

The analyte solution flow passes through the electrospray needle across which there is a high potential difference. This forces the spraying of charged droplets from the needle with a surface charge of the same polarity to the charge on the needle. The droplets are repelled from the needle and as the droplets traverse the space between the needle tip and the cone, solvent evaporation occurs. The droplet then shrinks until it reaches the point that the surface tension can no longer sustain the charge, at which point the droplet is fragmented into small molecules. These charged analyte molecules (they are not strictly ions) can be singly or multiply charged. Diagram illustrated by Dr. P Gates (University of Bristol).

Since the m/z value is in many cases not unique for a protein or a peptide, it is often not sufficient for an unambiguous protein identification (Frohlich & Arnold, 2006). Therefore, tandem mass spectrometry (MS/MS) techniques have been employed which consist of firstly identifying the m/z of the precursor ion and secondly, isolating the precursor ion from all other peptide ions in the instrument and dissociating them into fragments, the masses of which are determined in the second MS step (McLafferty, 1981). The most common fragmentation techniques for peptide ions is the collision induced dissociation (CID), where peptide ions collide with inert gas atoms or molecules such as He, Ar, N_2 (Hayes & Gross, 1990). The MS/MS spectra obtained are correlated with theoretical spectra calculated from sequence databases as described before.

Identification of proteins in samples which have been subject to mass spectrometry is carried out using a probability-based scoring algorithm, Mascot, that uses information from either molecular weights of digested proteins or tandem mass spectrometry (MS/MS) data against protein databases such as the national centre for biotechnology information (NCBI) or SwissProt (Perkins *et al*., 1999).

A summary of the potential advantages and disadvantages of different mass spectrometry or MS-linked techniques with regard to large-scale quantitation of proteins in complex mixtures is shown in **Table 1.2**.

Two main approaches have been suggested for *in vitro* analysis of proteins by mass spectrometry (Chait, 2006). One involves the use of intact proteins which have not been subject to cleavage and would therefore be a true representation of the state of the complex mixture *in vivo*. These methods are generally referred to as 'top-down' approaches. On the other hand many mass spectrometry techniques depend on labelling of the products of protein digestion. These methods have proven successful in relative quantitation of proteins in complex mixtures and are generally classified as 'bottom-up' approaches

Methodology	Mass range (kDa)	Advantages	Disadvantages
MALDI / SELDI TOF MS	> 950	Have been successful in identification of proteins in simple mixtures. Use in relative/absolute quantitation in complex mixtures is not established	Protein identification in complex mixtures can only be carried out with certainty when coupled with MS/MS.
ESI	> 100	Allows identification of proteins of up to 100 kDa in conjunction with LC-MS. Use in quantitation of proteins has been assessed in conjunction with FT-ICR (Hakansson *et al*., 2003).	Limited efficiency for complex mixtures.
CE	> 2-10	High sensitivity achieved when coupled with other instruments e.g. FT-ICR. Relative quantitation of proteins in urine has been successful (Weissenger *et al*., 2004). Low volume of starting material required (~10 nl of plasma).	Detection is limited to lower molecular weight fraction (<15 kDa).
Label-based e.g. iTRAQ	> 100	Relative quantitation of proteins in complex mixtures has been carried out reliably (DeSouza *et al*., 2005; Ziese, 2006) when coupled with MS/MS.	Involves extensive upstream sample treatment i.e. gel separation and immunodepletion. A major limitation is the expensive cost of reagents.

Table 1.2 Summary of the techniques used in identification and characterisation of proteins in complex mixtures.

A summary of approaches involved in mass spectrometry or MS-linked techniques used in the identification and/or quantitation of proteins in complex mixtures. Major advantages and disadvantages of each technique have also been summarised.

1.2.1.2.1 Top-down' approach

ESI has been the standard ionisation method for liquid chromatography (LC)-MS and LC- MS/MS (**Figure 1.5**), although separated protein fractions can also be deposited onto a MALDI target for analysis. In LC-ESI, the elute is directly pumped, at low flow rates (~ several nl/min) through a small diameter electrospray emitter needle thereby spraying the protein solution into the mass spectrometer (Righetti *et al*., 2005). ESI is especially useful in the analysis of macromolecules as it overcomes the propensity of these molecules to fragment when ionised (Baldwin, 2004). This method of ionisation can be employed by mass spectrometry techniques such as Fourier transform ion cyclotron resonance (FT-ICR) in the 'top-down' analysis of proteins (Adkins *et al*., 2005). FT-ICR MS is a high resolution technique with high accuracy in determining protein masses, with sub-femto mol sensitivity (Hipple *et al*., 1949). The technique is based on trapping ions in a fixed magnetic field and applying an excitation pulse which results in higher energy movement of ions, whose frequency is proportional to the mass-to-charge ratio.

MALDI-TOF MS has enabled the analysis of a large number of proteins and peptides (Hillenkamp *et al*., 1991). In contrast to ELISA, MALDI-TOF MS allows detection of the analyte directly by an intrinsic property – its molecular weight (**Figure 1.5**). An MS protein profile can be used to detect multiple biomarkers simultaneously, by the use of sophisticated pattern determination and class prediction algorithms. However, there have been concerns about the reproducibility and coverage of MALDI/SELDI-TOF MS. MALDI-TOF profiling of plasma proteins and peptides has been carried out after sample pre-treatment with magnetic beads (Villanueva *et al*., 2004; Villanueva *et al*., 2006). These studies showed that plasma from brain tumour patients could be distinguished from controls with 96% certainty based on a pattern of 274 peptide masses. A study of plasma samples from patients with ovarian cancer, compared with normal controls, observed protein profiles (1,000-12,000 kDa) using MALDI-TOF MS which significantly differed between cases and controls (Callesen *et al*., 2005). On the other hand, a study of differential protein patterns between plasma samples collected from 82 pancreatic cancer patients and 40 controls showed low specificity associated with detection of only the highly abundant plasma proteins (Koomen *et al*., 2005).

Protein fractionation prior to MS is a technique that can be used to solve the problem of sample complexity. Capillary electrophoresis coupled with FT-ICR followed by MS is an example of such techniques. In capillary electrophoresis (CE), a large voltage (usually around 30 kV) is applied across a narrow bore silica capillary (75 mm internal diameter) and separation is based on differences in solute velocity in an electric field. The solute velocity is a function of an ion's electrophoretic mobility and applied voltage. The bulk flow of molecules through the system, when a current is applied, is by electro-osmotic flow (EOF). The magnitude of EOF is strongly dependent on pH; at high pH, EOF is greater than at low pH. CE-MS/MS analysis requires only nanolitres of sample and each analysis can be completed in less than 60 minutes. Capillary electrophoretic techniques have been employed in protein quantitation of complex bodily mixtures including urine (Weissinger *et al*., 2004) and plasma (Gay-Bellile *et al*., 2003; Zinellu *et al*., 2005). However, to date, studies have only succeeded in quantitative assessment of highly abundant proteins such as creatinine, albumin and globulins in plasma, and fragments of albumin in urine. Furthermore, all quantitation procedures have involved comparison of disease cases and controls and no assessment of normal variation has been carried out.

1.2.1.2.2 Bottom-up' approach

Conventional methods of protein characterisation such as the TOF instruments, MALDI and SELDI, or ionisation techniques, often fail with low abundance proteins due to the sequestering effect of the high abundance proteins in complex mixtures. Quantitative analysis of plasma proteins is also limited to differential protein profiles compared between disease cases and healthy controls (Villanueva *et al*., 2004; Villanueva *et al*., 2006; Callesen *et al*., 2005) which has also been shown to suffer from lack of sensitivity (Koomen *et al*., 2005). Analysis is also limited to those proteins that fit the optimal size or concentration range for the MS methods (Koomen *et al*., 2005). To overcome some of the above limitations, a number of advances involving the stable isotope labelling technology for relative quantitative profiling via mass spectrometry have been established.

The relative quantitation approach by labelling is based on the abundance ratios of the individual proteins in two or more biological stages or samples (Frohlich

& Arnold, 2006). A common strategy is to firstly label the samples with chemically identical reagents (i.e. stable non-radioactive isotopes) differing in their isotopic composition and hence in their molecular weight (so called heavy and light reagents), and secondly to pool samples after labelling and thirdly, to perform LC-MS/MS experiments.

Isotope coded affinity tags (ICAT) are the most widely used of these techniques (Gygi *et al*., 1999), in which two samples (i.e. control versus experimental) are covalently labelled with chemically identical affinity tags (e.g. biotin) that differ only in isotopic composition. The samples are combined, enzymatically digested and the labelled peptides are selectively enriched by affinity chromatography based on the tags (e.g. biotin avidin). ICAT-labelled peptide fragments differ in mass by a known amount and can be separated and quantified by mass spectrometry. A major drawback of this technique is that only cysteine residues can be labelled, resulting in loss of many important proteins, especially those with post-translational modifications. ICAT has been successfully applied to the comparison of the relative quantities of proteins between two samples from cultured liver cells (Yan *et al*., 2004), animal tissues (Harris *et al*., 2003) and human plasma (Wu *et al*., 2003). A similar approach, referred to as stable isotope labelling by amino acids in cell culture, SILAC, involves labelling proteins by growing cells in media containing isotopically labelled amino acids including ^{2}H-leucine, ^{13}C-lysine, ^{13}C-tyrosine, ^{13}C-arginine and ^{13}C/^{15}N-arginine rather than using a covalently-linked tag (Ong *et al*., 2002). An advantage of SILAC is that the stable isotopic tags are incorporated into the early stages of sample preparation and thus reduce variation between samples. A disadvantage is that this method is not practical for analysing biological samples that cannot be grown in culture, such as bodily fluids.

A variation of the ICAT technology, which has been reliably used for relative quantitation of proteins in multiple samples involves the use of isobaric tags for relative and absolute quantitation (iTRAQ™) (Ross *et al*., 2004). The iTRAQ reagents consist of a reporter group, a balance group and a peptide reactive group. The peptide reactive group specifically reacts with primary amine groups of peptides. The reporter group is a tag with a mass of 114, 115, 116 and 117 Da depending on differential isotopic combinations in each individual reagent. The balance group

ranges in mass between 28 to 31 Da to ensure that combined mass of the reporter and the balance groups remain constant (145 Da). Peptide labels with different isotopes are isobaric and are chromatographically indistinguishable which are important for accurate quantitation. During collision induced dissociation, the reporter group ions fragment from the backbone peptides displaying distinct masses of 114 to 117 Da. The intensity of these fragments is used for quantitation of the individual representative peptides. The most important advantage of iTRAQ is that it allows labelling of up to four samples within a single experiment. This technology has been successfully applied to the quantitation of proteins in yeast (Ross *et al*., 2004) and also to the relative quantitation of proteins in tissue samples of diseased individuals compared with the normal controls (DeSouza *et al*., 2005).

Another method for separation and characterisation of protein molecules is ion exchange chromatography (IEC). IEC is based on the interaction between charged solute molecules and an oppositely charged resin covalently linked to a chromatography matrix (Bonnerjera *et al*., 1986). Following the reversible binding of the desired solute molecules to the matrix based on their net ionic charge, and release of the unbound substances, the bound fraction is eluted by changing the elution conditions to ones that are unfavourable for ionic bonding of the solute molecules. This is achieved by increasing the ionic strength of the eluting buffer or by changing its pH. IEC coupled with LC-MS/MS is a label-free method for relative quantitation of proteins and requires comparison of identical proteolytic peptides to accurately determine relative ratios of the proteins which depend on the accuracy of mass measurement and chromatographic reproducibility (Frohlich & Arnold, 2006). IEC LC-MS/MS has been successfully used to determine relative quantities of proteins in human plasma by the use of internal controls of known concentration (Chelius & Bondarenko, 2002).

1.3 Alzheimer disease

Alzheimer disease (OMIM # 104300) is the most common cause of dementia in Western societies. United Nation population projections estimate that the number of people older than 80 years will approach 370 million by the year 2050. Currently it

is estimated that 50% of people over the age of 85 are afflicted with Alzheimer disease. If these statistics hold true, in 50 years, more than 100 million people worldwide will suffer from dementia.

The onset of Alzheimer disease (AD) is gradual with the appearance of clinical symptoms occurring in most sufferers between 60-70 years of age. AD is characterised by progressive loss of cognitive function. The pathological hallmarks of the disease include formation of extracellular senile (neuritic) plaques containing β-amyloid (Aβ) and intracellular tau-rich neurofibrillary tangles (NFT), associated with widespread neuronal degeneration (Glenner & Wong, 1984; Masters *et al*., 1985; Spires & Hyman, 2004). Morphological changes preceding cell death change the connectivity of neurons and consequently disrupt the signalling in neuronal circuits. Neurons involved in higher-order functions such as learning, memory, consciousness, perception and self-awareness are particularly vulnerable in AD (Spires & Hyman, 2004). The assembly of Aβ into protofibrils, fibrils, or other toxic moieties has long been implicated as a key pathologic trigger of AD, leading to the amyloid hypothesis of Alzheimer disease (Hardy & Selkoe, 2002) (**Figure 1.7**). The cloning of the gene encoding a large single transmembrane protein, Amyloid Precursor Protein (*APP*) and discovery of mutations in familial Alzheimer disease (FAD) patients led to the hypothesis that Aβ accumulation is the primary event in AD pathogenesis (Hardy & Selkoe, 2002). Earlier recognition that trisomy 21 (Down's syndrome) invariably led to the neuropathology of AD by the 4^{th} or 5^{th} decade of life further reinforced the hypothesis since *APP* is localised to chromosome 21 (Glenner & Wong, 1984). APP is a type 1 transmembrane and cell surface receptor protein expressed in all cell types but mainly in the brain which is believed to be involved in synapse formation and transmission (Priller *et al*., 2006). Since amyloid plaques are a defining feature of AD neuropathology, and Aβ can be detected in cerebro-spinal fluid (CSF) and plasma, Aβ measures in biological fluids are compelling candidate biomarkers for AD.

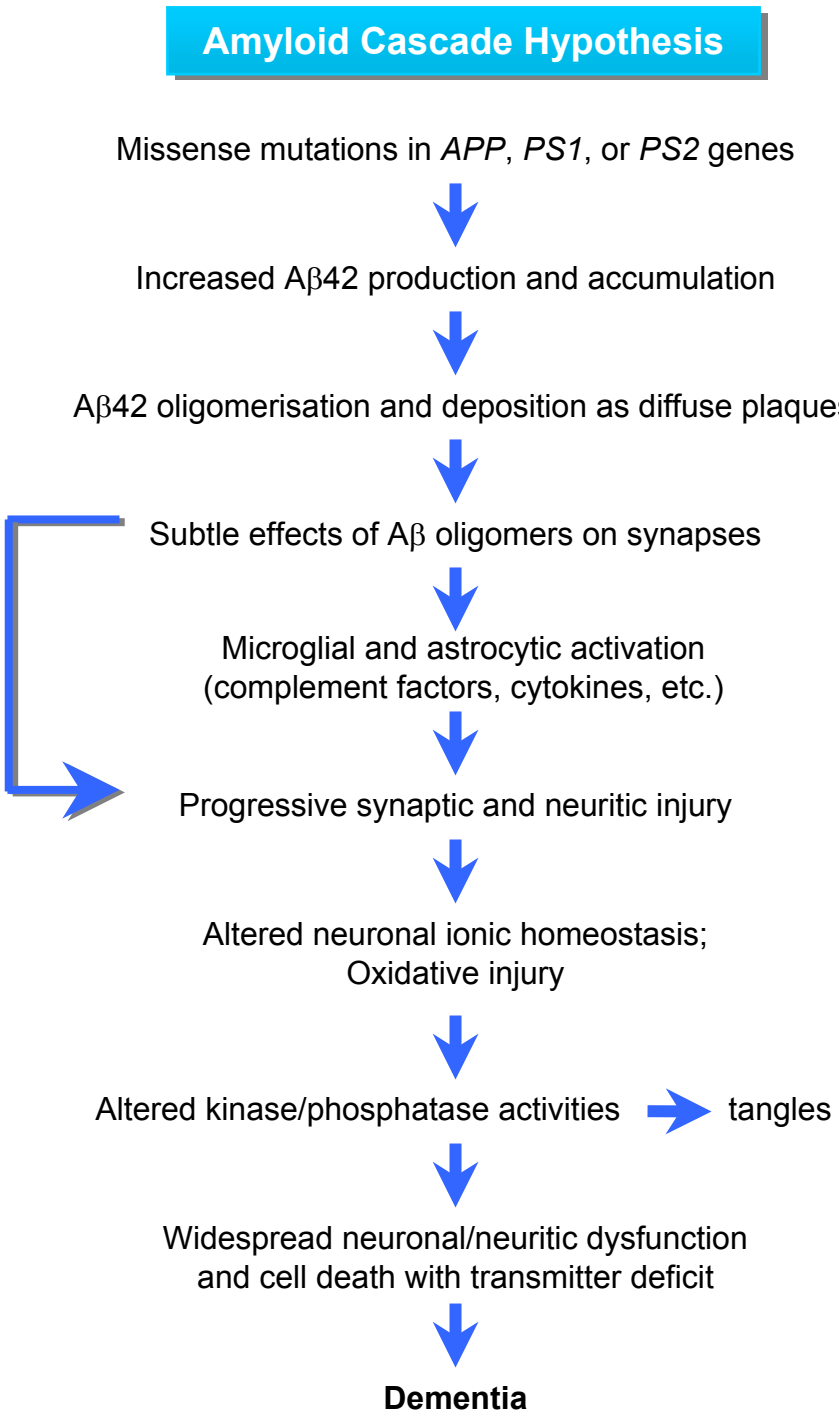


Figure 1.7 The sequence of pathogenic events leading to Alzheimer disease as proposed in the amyloid cascade hypothesis.

The curved blue arrow indicates that Aβ oligomers may directly injure the synapses and neurites of brain neurons, in addition to activating microglia and astrocytes. This phenomenon has been implicated as the key patho-physiological hallmark of AD. Diagram reproduced from Hardy and Selkoe (2002).

Aβ is a normal product of APP metabolism throughout life. In addition to causing FAD, mutations in the *APP* gene can also cause hereditary cerebral hemorrhage with amyloidosis (Dutch type), which have been shown to result in Aβ deposition largely in blood vessels and outside the brain parenchyma (van Broeckhoven *et al*., 1990). Aβ is proteolytically derived from APP as a result of sequential cleavages by β- and γ-secretases (Vetrivel & Thinakaran, 2006) (**Figure 1.8**). β-Secretase has been identified as a type I membrane aspartyl protease. γ-Secretase is believed to be a multi-protein complex composed of presenilins I and II, nicastrins, Anterior Pharynx defective 1 Homolog (*C. elegans*) (APH1), and Presenilin Enhancer 2 homolog (*C. elegans*) (PSEN), with presenilin believed to be the functional catalytic subunit. The C-termini of secreted N-terminal Aβ peptides are generated by three major intra-membranous cleavages, including one referred to as γ-cleavage, which cleaves at residues 40 or 42 for Aβ40 and Aβ42, respectively (**Figure 1.8**). This produces the C-termini of Aβ species that end at amino acids 40 (Aβ40) or 42 (Aβ42) with molecular masses of 4.3 kDa and 4.5 kDa, respectively. A second type of cleavage, ε-cleavage, occurs between Aβ residues 49 and 50 and produces the N-terminus of most of the APP intracellular domain (AICD). A third major cleavage site, ζ-cleavage, has also been identified at Aβ46, between the known γ- and ε-cleavage sites (Zhao *et al*., 2004). Aβ46 has recently been identified as an intermediate precursor of Aβ which is tightly associated with presenilin in intact cells (Zhao *et al*., 2005). Zhao and colleagues showed that another intermediate product, Aβ49, is generated by ε-cleavage, which is further processed by ζ- and γ-cleavage to generate Aβ46 and ultimately the secreted Aβ40 and Aβ42.

Amyloid fibrils are filamentous structures with width of ~ 10 nm and a length of 0.1-10 μm. A defining feature, originally revealed by X-ray fibre diffraction analysis is the presence of cross-β structures (Chaney *et al*., 1998). Ribbon-like β-sheets are formed by β-strands running nearly perpendicular to the long axis of the fibril and hydrogen bonds that run nearly parallel to the long axis (Ross & Poirier, 2004).

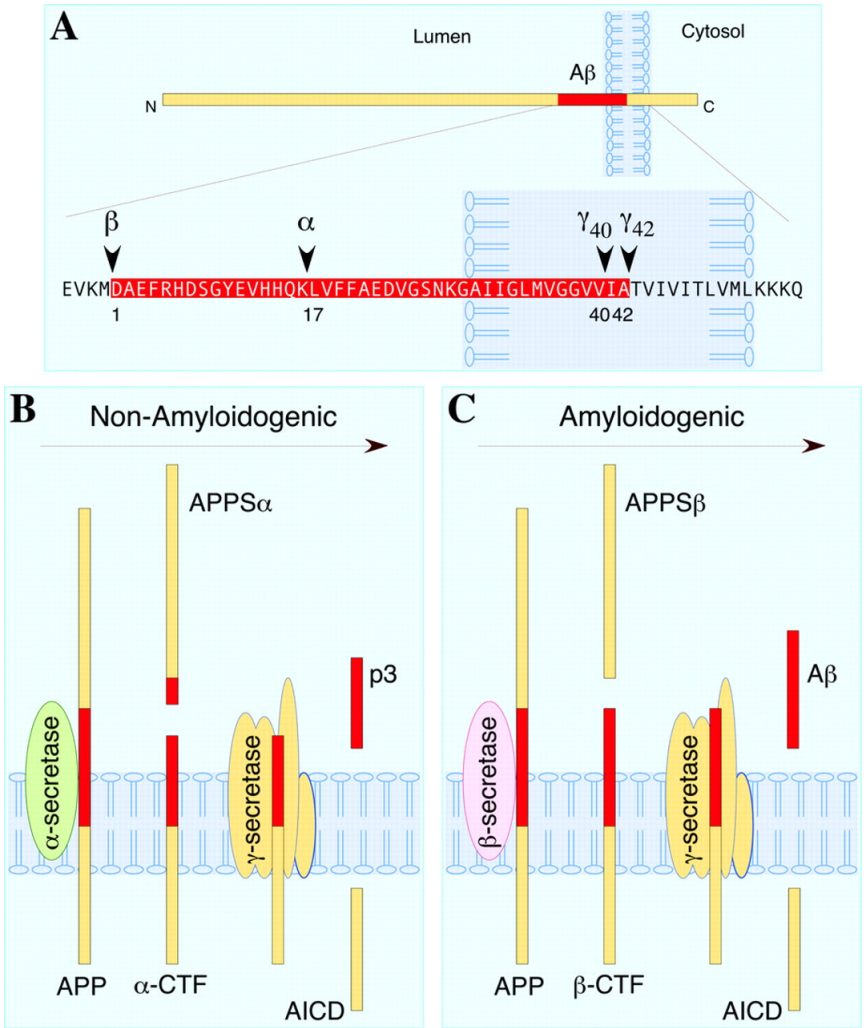


Figure 1.8 Proteolytic processing of APP.

A. Schematic structure of APP is shown with the Aβ domain shaded in red and enlarged. The sites of cleavage by α-, β-, and γ-secretases are indicated along with Aβ numbering from the N-terminus of Aβ. **B.** Non-amyloidogenic processing of APP refers to sequential processing of APP by membrane-bound α- and γ-secretases. α-Secretase cleaves within the Aβ domain, thus precluding generation of intact Aβ peptide. The fates of N-terminally truncated Aβ (p3) and APP intracellular domain (AICD) is not fully resolved. **C.** Amyloidogenic processing of APP is carried out by sequential action of membrane-bound β- and γ-secretases. The N-terminal of APP is extracellular while the C-terminal is intracellular. Diagram adapted from Kulandaivelu *et al*., 2006.

Covalent modification of proteins such as oxidative modifications associated with ageing may also facilitate aggregation, as in sporadic forms of neurodegenerative diseases (Ross & Poirier, 2004). Ageing may also decrease the ability of the cell to clear misfolded proteins (Ross & Poirier, 2004).

1.3.1 Genetics of Alzheimer disease

The initiation of aggregation can be due to protein-coding mutations, which result in the autosomal dominant FAD, which is believed to account for ~1% of all AD cases (St George-Hyslop *et al*., 1987; Finckh *et al*., 2005). The age at onset in these families is typically in the 40's and 50's, although sometimes as early as the 20's. Pathogenic FAD mutations usually cluster at, or very near, the sites within APP that are normally cleaved by α-, β- and γ-secretases. These mutations promote the generation of Aβ by favouring proteolytic processing of APP by β- or γ-secretase. Other single gene mutations have also been identified in *PSEN1* (chromosome 14) and *PSEN2* (chromosome 1), the genes encoding presenilin 1 and 2, respectively (Rubinsztein, 1997; Walker *et al*., 2005). Furthermore, *APP* mutations internal to the Aβ sequence heighten the self-aggregation of Aβ into amyloid fibrils (Hardy & Selkoe, 2002) (**Figure 1.8**).

Alzheimer disease provides an excellent paradigm for the genetic basis of a complex disorder, with contributions from both common modifier genes and rare variants of large effect. Heritability of AD has been estimated to be around 67% and 22% in monozygotic and dizygotic twins, respectively (Gatz *et al.*, 1997), suggesting that genetic variation plays a significant role in the disease process. However, the major insights into disease mechanisms to date have come from mutations in genes that are so rare that they make essentially no contribution to the heritability of the disease as a whole (Chinnery, 2006). Case-control studies have high sensitivity for detecting risk alleles influencing multifactorial disorders; however, almost none of the association findings in late onset AD (LOAD) have turned out to be reproducible and therefore most fall into the category of false-positive results (Schellenberg *et al*., 2004; Prince *et al*., 2001; Finckh, 2003).

To date only one gene, *APOE*, has been reliably associated with LOAD in multiple studies. The *APOE* gene, described later in more detail, has three common

alleles *APOE*E2*, *APOE*E3* and *APOE*E4*, and serves as one of the best examples of the common disease-common variant hypothesis (Reich & Lander, 2001). The *APOE*E4* allele has been associated with an increased risk of AD in many studies, by promoting an earlier age at onset in a dosage-dependant manor (Saunders *et al*., 1993; Corder *et al*., 1993). For example, a person of European descent who is homozygous for the *APOE*E4* allele has a lifetime risk of AD that is two times as great as that for a heterozygous individual and four times as great as that for someone of European descent who does not have an *E4* allele (Bird, 2005). Although the association between the *E4* allele and AD has been replicated in many studies, the protective effect of the less common *E2* allele is still disputed and requires further investigation (Corder *et al*., 1994; Smith *et al*., 1994). The estimated contribution of *APOE*E4* to the total variation in age at onset of AD has been estimated to be 7-9% (Daw *et al*., 2000). Daw and colleagues suggest that several other loci may play a significantly larger role than *APOE* in susceptibility to LOAD (Daw *et al*., 2000).

The analysis of intermediate or quantitative traits such as plasma or CSF Aβ is a major advantage in that it provides greater power for detecting genetic influences on AD compared with binary case/control studies (Brookes & Prince, 2005).

1.3.2 Biochemistry and genetics of amyloid-β

Secreted, soluble Aβ40 and Aβ42 are products of normal cell metabolism which are found in various bodily fluids, including plasma and CSF. In AD brains, Aβ42 has been found to be deposited first and constitutes the predominant form in senile plaques (Hardy & Selkoe, 2002). Aβ40 is deposited later in the disease and is more prominent in vascular amyloid deposits. Although cerebral deposition of Aβ42 is unlikely to directly result from increased plasma Aβ42, elevated plasma levels probably do reflect a generalised alteration in Aβ in both neural and non-neural cells (Mayeux *et al*., 1999). It has been proposed that elevated plasma levels of the released Aβ42 may be detected several years prior to the onset of symptoms (**Figure 1.9**) (Mayeux *et al*., 2003) and Aβ42 concentrations are believed to gradually decrease in CSF and plasma because of preferential sequestration as insoluble deposits in brain (**Figure 1.10**) (van Oijen *et al*., 2006).

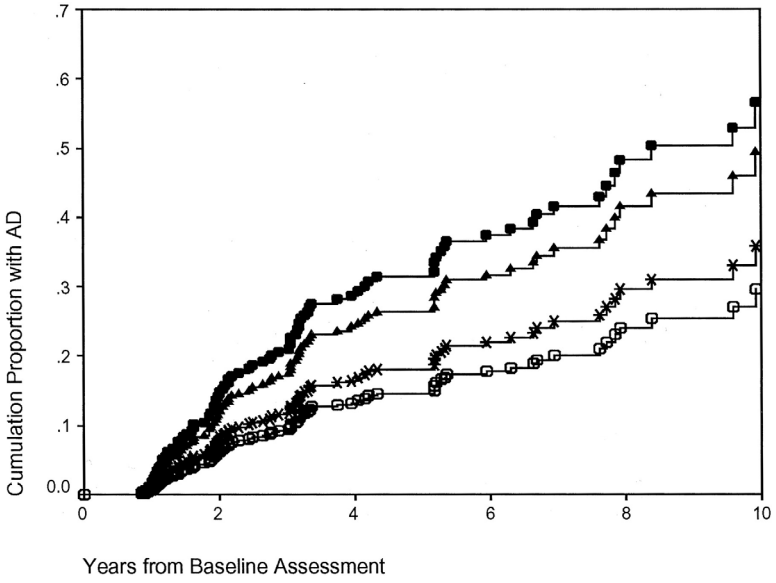


Figure 1.9 Risk of Alzheimer disease by quartiles of plasma Aβ42 concentration.

The graph shows the cumulative proportion of individuals developing Alzheimer disease (AD) over a ten year period of follow-up by the plasma Aβ42 concentrations at the baseline (mean age=83.2 years). These results show that individuals with increased plasma Aβ42 concentrations are at an increased risk of developing AD. Filled circles, Aβ42>84.15 pg/ml; triangles, 60.2<Aβ42<84.15 pg/ml; asterisks, 36.97<Aβ42<60.2 pg/ml; open circles, Aβ42<36.97. Figure adapted from Mayeux *et al*., 2003.

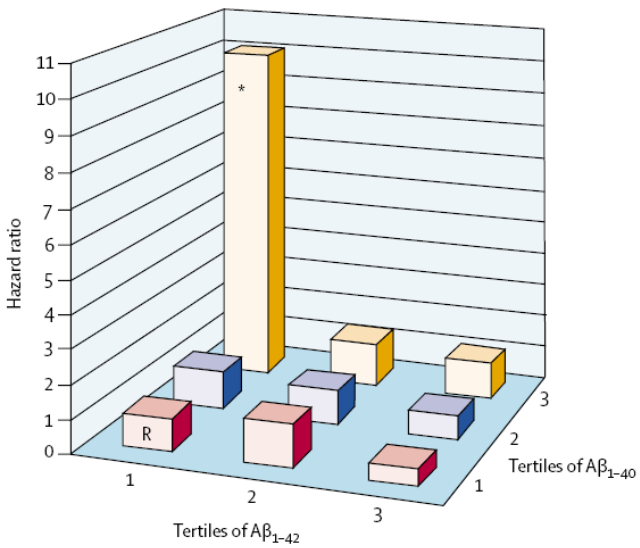


Figure 1.10 Hazard ratios for dementia by concentrations of Aβ40 and Aβ42.

Diagram showing the significant effect of high plasma concentrations of Aβ40 combined with low concentrations of Aβ42 on increased risk of dementia. These results show that individuals with concentrations in the highest tertile of Aβ40 combined with concentrations in the lowest tertile of Aβ42 compared with individuals with concentrations in the lowest tertiles of both Aβ40 and Aβ42 (R), are at 10 times increased risk of developing AD. *P < 0.0001. Diagram adapted from van Oijen *et al*., 2006.

The genetic component responsible for the variability of plasma Aβ has been estimated previously from the additive (narrow-sense) heritability of plasma Aβ40 and Aβ42 (Ertekin-Taner *et al*., 2001). This study was based on analysis of 15 extended LOAD families (N = 396) ascertained via AD patients or first-degree relatives with high plasma Aβ levels. Narrow-sense heritability was found to be 73% for Aβ42 and 54% for Aβ40 (both highly significant) using a variance component methodology.

Ertekin-Taner and colleagues have since analysed two genes, Catenin Cadherin-associated protein alpha 3 (*CTNNA3*) and Plasminogen Activator Urokinase (*PLAU*)) on chromosome 10q, previously shown to be linked to AD using extended LOAD pedigrees (Ertekin-Taner *et al*., 2000). *CTNNA3* encodes α-T catenin, which is expressed at high levels in human and mouse brain and interacts with β-catenin. This makes *CTNNA3* a strong functional candidate because β-catenin interacts with presenilin 1 (Zhang *et al*., 1998), in which mutations have been shown to increase Aβ42 concentration in plasma and cause early-onset FAD (St George-Hyslop *et al*., 1987).

PLAU, a gene encoding urokinase-type plasminogen activator which converts plasminogen to plasmin, has also been shown to be associated with risk of LOAD in 10 extended families (N = 554) ascertained via the presence of a proband with high plasma Aβ concentration (Ertekin-Taner *et al*., 2005). This association has been replicated in six independent case/control LOAD cohorts (N_{each} > 200) (Ertekin-Taner *et al*., 2005). Others have reported the absence of association between variants in *PLAU* and risk of LOAD in three independent cohorts, consisting of 200-800 individuals (Myers *et al*., 2004). The PLAU protein is a serine protease which is involved in degradation of the extracellular matrix. Aβ aggregates might induce *PLAU* expression, thereby increasing plasmin which degrades both aggregated and non-aggregated forms of Aβ (Ertekin-Taner *et al*., 2003).

Furthermore, plasma Aβ40 and Aβ42 have also been shown to increase in *PLAU* knockout (KO) mice by 11 months of age, suggesting that this effect may become more pronounced with ageing (Ertekin-Taner *et al*., 2005).

The above studies have reinforced the need to identify genetic variants influencing both susceptibility to LOAD and circulating Aβ concentrations. A thorough analysis of QTL associated with changes in plasma Aβ concentrations could lead to the discovery of new biochemical pathways involved in the pathogenesis of Alzheimer disease.

1.3.3 Rodent models of Alzheimer disease

Several rodent models of AD have been developed which display the amyloid plaque pathology (as a result of *APP* over-expression), the neurofibrillary tangle (NFT) pathology (as a result of microtubule-associated protein tau tangle formation) or both.

Table 1.3 summarises the models developed to date and their advantages and disadvantages in relation to the neuropathology of plaque or NFT phenotypes. Other phenotypes such as memory deficits and age of onset of pathology have also been assessed in these models (Spires & Hyman, 2005).

The amyloid plaque models have aimed at over-expressing human *APP* or fragments of it that contain missense variants which have been detected in the familial forms of the disease in humans. Although all such models have resulted in plaque formation as early as three month, some, like the PDAPP mice (Games *et al*., 1995), do not display global neuronal loss (Irizarry *et al*., 1997). The Tg2576 (Hsiao *et al*., 1996), APP23 (Sturchler-Pierrat *et al*., 1997), TgCRND8 (Dudal *et al*., 2004) and APPSwe TgC3-3 (Borchelt *et al*., 1997) mice all develop the amyloid plaque neuropathology, with neural loss of up to 14% (in the case of APP23) and memory loss (not reported for APPSwe TgC3-3) between 3 and 18 months (**Table 1.3**).

Mutations in the human *APP* gene cause only a small fraction of the already rare number of FAD cases, therefore, FAD-associated presenilin variants have been used to generate animal models carrying both *APP* and *PSEN* missense variants. For example, the PSAPP mice (cross between Tg2576 and *PSEN1* M146L) (Holcomb *et al.*, 1998) show elevated concentrations of Aβ42 in the brain and an accelerated rate of Aβ deposition in the brain compared with Tg2576 mice (**Table 1.3**).

Name	Animal model	Gene(s) overexpressed	Neuropathology plaques	P-tau immunoreactivity	NFT	Memory deficit	Age at onset (months)
PDAPP	Mouse	*APP* V717F mutation	Yes	Yes	No	Yes	6 - 8
Tg2576	Mouse	*APP* (Swe) cDNA	Yes	Yes	No	Yes	9 - 11
APP23	Mouse	*APP* (Swe) cDNA	Yes	Yes	No	Yes	6
TgCRND8	Mouse	*APP* (Swe) cDNA (V717F) mutation	Yes	n.a.	No	Yes	3
APPSwe TgC3-3	Mouse	*APP* cDNA	Yes	n.a.	n.a.	n.a.	18
PSAPP	Mouse	Tg2576 and *PSEN* M146L	Yes	Yes	n.a.	Yes	6
Tg478/1116/11587	Rat	*APP* (Swe), *APP* V717F, *PSEN1* M146V	Yes	n.a.	n.a.	n.a.	9
ALZ7	Mouse	*MAPT* (tau)	No	Yes	No	n.a.	-
ALZ17	Mouse	*MAPT* (tau)	No	Yes	No	n.a.	-
3xTg-AD	Mouse	*APP* (Swe), *PSEN1* (M146V), *MAPT* (P301L)	Yes	Yes	Yes	n.a.	3

Table 1.3 Transgenic rodent models of Alzheimer disease pathology

Summary of the transgenic rodent models showing the Alzheimer disease pathology as a result of over-expression (using neuron-specific promoters) of the genes implicated in the human familial form of AD. P-tau, phosphorylated tau; NFT, neurofibrillary tangle; n.a., not available. Table adapted from Spires & Hyman, 2005).

The Tg478/1116/11587 transgenic rat model (carrying two *APP* variants and *PSEN1* M146V) (Spires & Hyman, 2005) shows amyloid deposition similar to that seen in Tg2576 mice (**Table 1.3**). The advantage of using rat models of AD is that several types of experiment such as behavioural testing of cognitive function and electrophysiological recordings can be carried out more easily than in mice.

Several mouse models of the NFT phenotype have also been developed. In healthy adult brain, MAPT is located in axons but in AD and other taupathies (such as Parkinson's disease), it is hyperphosphorylated and is found in cell bodies and dendrites as well as axons (Brion *et al*., 1985). The two TNF mouse models, ALZ7 (Gotz *et al*., 1995) and ALZ17 (Probst *et al*., 2000), show hyperphosphorylation of MAPT but do not develop NFT (**Table 1.3**). These two models show that overexpression of the normal human MAPT is insufficient to induce NFT formation but the localisation of MAPT in neurons does resemble the pre-tangle state in AD. The best model developed to date, is the 3xTg-AD mouse (Oddo *et al*., 2003) which harbours the *APP* (Swedish), *PSEN1* (M146V) and *MAPT* (P301L), which develops Aβ plaques around 3 months of age and tangles after the amyloid pathology. The regional and temporal development of pathology closely mimics the development of pathology in AD, making this a good disease model.

The development of animal models of AD has led to promising targets for therapeutics. Successful findings include the targeting of the PSEN1 action by nonsteroidal anti-inflammatory drugs (NSAIDs) which modulate γ-secretase cleavage in 11-month old Tg2576 mice leading to reduced plaque formation (Yan *et al*., 2003). Metal ions have been implicated as Aβ-interacting molecules leading to reduced risk of AD (Gnjec *et al*., 2002), and treatment of Tg2576 mice with the antibiotic clioquinol that chelates zinc and copper ions has resulted in decreased Aβ brain deposition (Lee *et al*., 2002). Finally, immunotherapy with Aβ antibodies in 3xTg mice has resulted in plaque clearance and a reduction in NFT formation (Oddo *et al*., 2004). However, the clinical trial involving human AD patients was unsuccessful due to 6% of the participants developing meningoencephalitis (Orgogozo *et al*., 2003).

1.4 Biochemistry and genetics of APOE

Apolipoprotein E (OMIM # 107741) is a member of the apolipoprotein multi-gene family, which also includes *APOA1*, *APOA2*, *APOA4*, *APOC1*, *APOC2* and *APOC3*. The *APOE* gene is located on chromosome 19q13.2 and is closely linked to the *APOC1/C2* gene complex. It consists of four exons and three introns spanning 3,597 nucleotides and encodes a 299 amino acid polypeptide (Eichner *et al*., 2002). The APOE protein is synthesised primarily in the liver but other organs and tissues, including brain, spleen, kidneys, gonads, adrenals and macrophages, also produce it. APOE plays a central role in plasma lipoprotein metabolism and in lipid transport within tissues (Mahley, 1988).

Three major APOE isoforms are found in humans, E2, E3 and E4 identified by isoelectric focusing (Utermann *et al*., 1979). These isoforms differ in amino acid sequences at two sites, residue 112 and residue 158. E2, E3 and E4 isoforms contain cys^{112}/cys^{158}, cys^{112}/arg^{158}, and arg^{112}/arg^{158}, respectively (Weisgraber *et al*., 1981; Rall *et al*., 1982).

The products of the three common *APOE* alleles differ in their affinity for binding low-density lipoprotein (LDL) receptors and lipoprotein particles (Weisgraber *et al*., 1981). The APOE isoforms interact differently with specific lipoprotein receptors, altering circulating levels of cholesterol (Dong *et al*., 1994). Studies have associated *APOE*E2* carriers with lower levels of plasma total and LDL cholesterol, compared with those in *APOE*E3* homozygotes (Schneider *et al*., 1981; Weisgraber *et al*., 1982; Utermann *et al*., 1985; Sing & Davignon, 1985). However, these results are open to dispute due to the low frequency of the *APOE*E2* allele in the general population (0.08 in the UK population) (Corbo *et al*., 1995; Lucotte *et al*., 1997). Moreover, few studies have large enough sample sizes to examine the less prevalent *E2*/*E2* genotype (0.003 in the UK population) (Corbo *et al*., 1995; Lucotte *et al*., 1997). The *APOE*E4* allele displays an opposite pattern with higher associated concentrations of total and LDL cholesterol (Utermann *et al*., 1985; Sing & Davignon, 1985; Davignon *et al*., 1988) and has been shown to be associated with increased risk of cardiovascular disease (Scuteri *et al*., 2005; Kang *et al*., 2005).

Studies of non-human primates have revealed that they carry the *APOE*E4* allele (Hanlon & Rubinsztein, 1995). This suggests that *E4* is the ancestral allele and that after the human and chimpanzee lineages split, *E3* and *E2* arose by mutation events and later spread among humans. Selection may also have contributed to the present distribution of the *APOE* alleles since the three alleles differ in their functional properties (Hanlon & Rubinsztein, 1995). The *E3* allele has been shown to be the most frequent in all human populations (0.78 in the UK population). Furthermore, the frequency of the *E3* allele is negatively correlated with that of *E4*. The highest frequencies of the *E3* allele are found in populations with a long established agricultural economy (Mediterranean, 0.85-0.89; or East Asian 0.82-0.87) compared with African populations (Pygmies, 0.54). The frequency of the *E4* allele remains higher in populations with lower plasma cholesterol levels (e.g. Pygmies, 0.41) than those observed in Western countries (Mediterranean, 0.07) in which an economy of foraging still exists, or the food supply has been scarce (Corbo *et al*., 1999).

1.4.1 APOE and dementia

Gene dosage of the *APOE*E4* allele has been shown to be a major risk factor for familial and sporadic Alzheimer disease (Saunders *et al*., 1993; Corder *et al*., 1993). The first reported association between *APOE*E4* and late-onset AD (LOAD) showed a reduction of 17 years in age of onset associated with every copy of the E4 allele (Corder *et al*., 1993). However, this may have been overestimated due to the small sample size used ($N < 100$) leading to the "Beavis effect" (Beavis, 1994). The Beavis effect was first demonstrated using large-scale simulation experiments, which examined 10 or 40 associations where each associated genetic marker explained a proportion of the phenotypic variance ranging from 0.75 to 9.5% with sample sizes ranging from 100 to 1,000. The experiment showed that the average estimates of phenotypic variances associated with correctly identified markers were greatly over-estimated if only 100 progeny were evaluated, slightly over-estimated if 500 progeny were evaluated and fairly close to the actual magnitude if 1,000 progeny were evaluated.

A study consisting of 115 unrelated LOAD cases and 243 healthy age-matched controls showed that *APOE*E4* was associated with increased risk of LOAD with odds ratios of 19.3 (two copies) and 4.4 (one copy) in an age-dependent manor (Corder *et al*., 1994). Similar findings have been obtained by several well-powered studies ($N > 400$) (Lucotte *et al*., 1994; Myers *et al*., 1996; Bickeboller *et al*., 1997).

APOE is produced in abundance in the brain and is induced to reach high concentrations in response to peripheral nerve injury (Mahley & Rall, 2000). It also appears to play a key role in repair by redistributing lipids to regenerating exons and to Schwann cells during remyelination (Mahley, 1988). A key to understanding the role of APOE in neurological diseases is believed to reside in determining how APOE modulates neural repair and protection (Mahley & Rall, 2000). It has been hypothesised that the APOE isoforms differentially affect amyloid plaque formation by direct protein-protein interaction, Aβ peptide metabolism, or both. Lipid-free APOE*E4 has been shown to form a stable complex with Aβ peptide *in vitro* (Strittmatter *et al*., 1993). In contrast, lipidated APOE*E3 has been found to bind Aβ peptide with a 20-fold higher affinity than lipidated E4 (**Figure 1.11**) (Stratman *et al*., 2005). The increased binding of lipidated E3 and E2 to Aβ species could enhance their clearance, preventing the accumulation of the neurotoxic Aβ species (Mahley & Huang, 1999).

1.5 Age-related macular degeneration

Age-related macular degeneration (AMD) (OMIM # 603075) is the most common cause of irreversible visual loss in the elderly population of developed countries, with approximately 25 million people affected worldwide (Evans & Wormald, 1996). AMD is characterised by degeneration of the macula. An intermediate stage of the disease, called age-related maculopathy (ARM), affects about 12 million individuals worldwide while the late-onset, advanced forms of the disease including dry AMD (geographic atrophy) and wet (neovascular) form of AMD or affects about 8 million people worldwide (Vingerling *et al*., 1995).

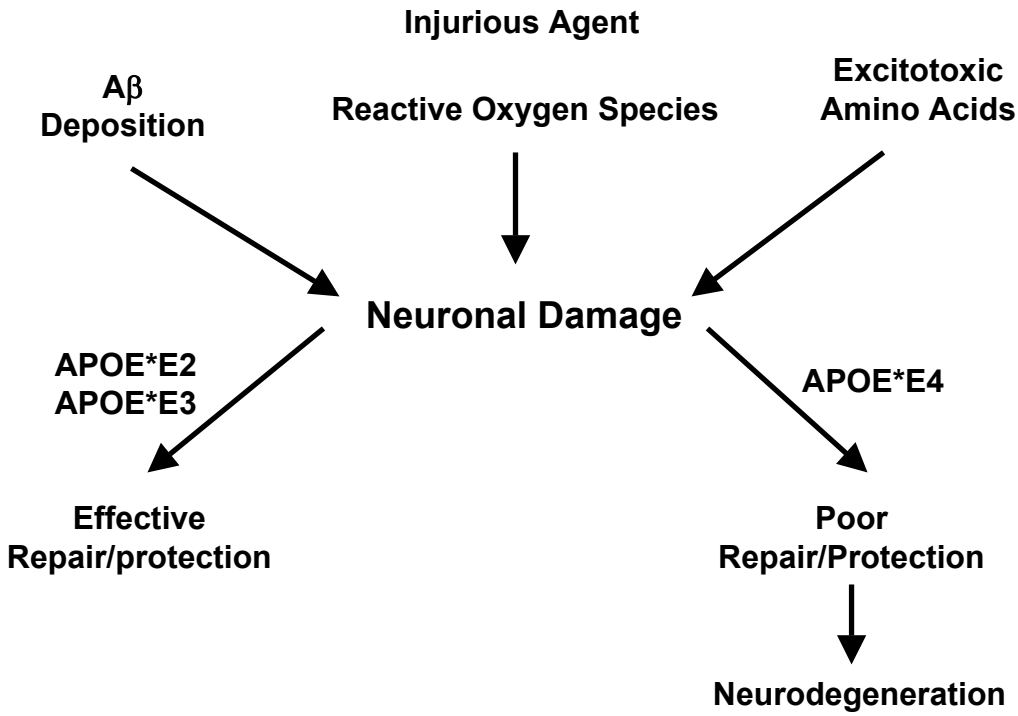


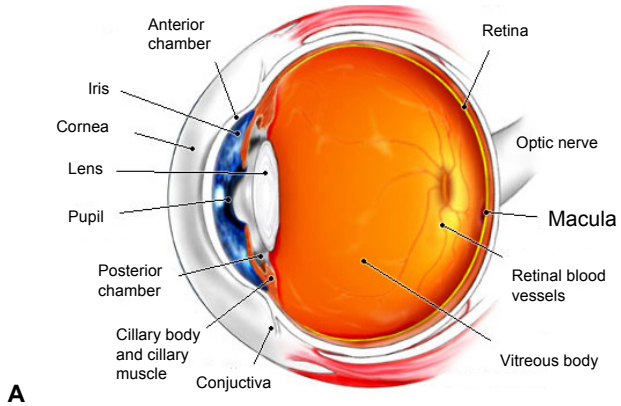
Figure 1.11 The hypothetical isoform-specific effects of APOE on neuronal repair, remodelling and protection.

Diagram adapted from Mahley and Rall, 2000.

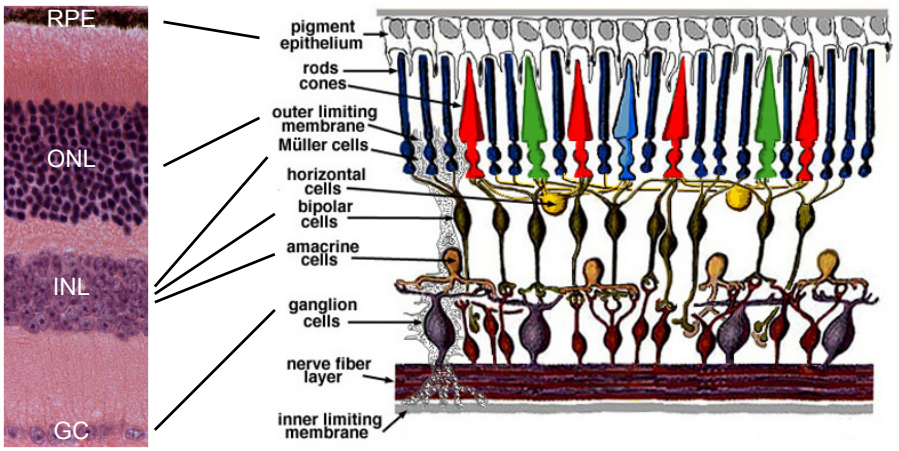
The wall of the human eyeball consists of three layers of tissue: the outer supporting layer (forming the tough outer layer or sclera and the transparent cornea), the vascular middle layer (containing choroidal blood vessels) and the inner layer consisting of neural retina and the retinal pigment epithelium (RPE) (**Figure 1.12**). The RPE cells support the photoreceptors (rods and cones) by engulfing and phagocytosis of up to 10% of the photoreceptor outer segment each day, maintenance of the Bruch's membrane (a specialised extracellular structure in the basement membrane of RPE), absorption of scattered light and transport of ions and fluids from the choriocapillaries (Lotery & Trump, 2007). The retina is a ten-layered tissue consisting of three layers of neurones including 1- the light-sensitive photoreceptor rods and cones, 2- horizontal, bipolar and amacrine cells, and 3- the ganglion cells. AMD affects the macular region of the eye defined as the central area of the retina (central 10%, 5 mm diameter) containing the fovea, the only region of the retina where the density of photoreceptors is sufficient to permit high acuity vision.

Early AMD is associated with moderate vision loss associated with extracellular deposits between the RPE and the Bruch's membrane (BrM), which can either be focal (drusen) or diffuse (basal deposits). Drusen are the hallmark of the disease and form between the basement membrane of the RPE and BrM. Drusen have been shown to contain a complex mix of proteins, lipids and lipoproteins including complement factors (C1q, C3a and C5a) (Nozaki *et al*., 2006), complement regulators (complement factor H, C-reactive protein and clusterin) (Hageman *et al*., 1999; Johnson *et al*., 2001), immunoglobulins (Johnson *et al*., 2001), amyloid-β, apolipoproteins and cholesterol esters (Mullins *et al*., 2000).

Age-related macular degeneration can be referred to as either "dry" (85% of cases) (geographic atrophy) or "wet" (neovascular or exudative) (15% of cases). The dry form is characterised by areas of hypopigmentation or depigmentation or apparent absence of the RPE, in which the choroidal vessels are more visible than in surrounding areas (Lotery & Trump, 2007). Wet AMD is characterised by the presence of choroidal neovascularisation (CNV) resulting in sub-retinal haemorrhage, epithelial detachment and severe loss of vision.



A



B

Figure 1.12 The structure of the human eye.

A. Schematic cross section of the human eye indicating its main structures. The macula is a cone-rich area located in the centre of the retina. **B.** Horizontal section of a human eye stained with haematoxylin and eosin (left) and a schematic diagram (right) showing the variety and the organisation of the retinal neurons. The outer nuclear layer (ONL) comprises the nuclei of rod and cone photoreceptor cells, where the phototransduction takes place. The inner nuclear layer (INL) contains the nuclei of Muller, horizontal, bipolar and amacrine cels. The ganglion cell (GC) layer is the most apical layer of the retina. The retina is in close proximity to the retinal pigment epithelium. Images taken from http://images.medicinenet.com/images/illustrations (A), Vlachantoni, 2006 (B, left) and http://webvision.med.utah.edu/imageswv/scheme.jpg (B, right).

AMD is a progressive disease with genetic influences as well as other strong risk factors including increasing age (Klein *et al*., 1992, 1997; Friedman *et al*., 2004), cigarette smoking (two- to four-fold increase in risk) (Christen *et al*., 1996; Seddon *et al*., 1996), diet (positive association between total fat intake and ARM) (Seddon *et al*., 1994, 2003a); Mares-Perlman *et al*., 1995) and BMI (high BMI associated with both dry and wet AMD) (Seddon *et al*., 2003b). A direct molecular link between oxidative damage to the macular RPE and AMD has been established by finding protein modifications in drusen and BrM that can be generated from the oxidation of lipids and carbohydrates (Hollyfield *et al*., 2003). Such modifications are likely to be found in drusen and BrM due to the photo-oxidative environment in the retina and the lipid-rich photoreceptor outer segments, which provide an excellent source of reactive oxygen species and oxidation products. It has also been suggested that immune-mediated complement activation triggered by unknown signals from the RPE, are causally involved in the process of drusen formation which leads to AMD (Hollyfield *et al*., 2003).

1.5.1 Genetics of AMD

Familial aggregation studies (comparing AMD prevalence in relatives of cases to AMD prevalence in relatives of controls) and twin studies have provided strong evidence of an underlying genetic component to AMD.

In familial aggregation studies, the prevalence of AMD in first-degree relatives of 119 AMD cases was found to be significantly higher than in relatives of 72 sex- and age-matched controls (23.7% vs. 11.6%) (Seddon *et al*., 1997). Similar results have been found by others using siblings and offspring of AMD cases compared with age- and sex-matched controls, showing that siblings of individuals with AMD have a three to six-fold increase in disease risk (Klaver *et al*., 1998b; Klein *et al*., 2001).

Twin studies have provided more direct evidence of AMD heritability by comparing disease concordance rates in monozygotic (MZ) versus dizygotic (DZ) twins. A study of 840 US elderly male twins (210 MZ and 181 DZ twin pairs) and 58 singletons, of whom 268 twin pairs showed signs of maculopathy (106 with advanced disease), found heritabilities of 0.67 for intermediate disease and 0.71 for

advanced disease (including geographic atrophy and exudative AMD) (Seddon *et al*., 2005). Similarly, a study of 506 pairs of female twins (226 MZ and 280 DZ pairs) with ARM (defined as the presence of soft drusen, > 63 μm in diameter, no late AMD) found the concordance for ARM to be 0.37 (in MZ twins) and 0.19 (in DZ twins) and the heritability of ARM was estimated to be 0.45 (Hammond *et al*., 2002).

Population-based studies of incidence and prevalence of AMD have shown that white individuals of European ancestry in US have a higher age-adjusted risk of developing AMD compared with individuals from other ethnic groups (Friedman *et al*., 2004). A lower prevalence of AMD has also been reported in populations from Japan and China compared with Europe (Oshima *et al*., 2001; Li *et al*., 2006b). It is not clear if these population-specific differences are due to genetic or environmental factors, or both; however, differences in disease prevalence between different ethnic groups emphasise the importance in design of genetic studies (Pritchard *et al*., 1999).

Nearly every chromosome in the human genome has been implicated by one or more genome-wide linkage studies for AMD, indicating that multiple genes may by involved in the development of AMD. However, most replicated linkage findings with statistical significance (LOD ≥ 3) have been on chromosomes 1q25-31 (Klein *et al*., 1998; Seddon *et al*., 2003c; Majewski *et al*., 2003; Iyengar *et al*., 2004; Jakobsdottir *et al*., 2005; Santangelo *et al*., 2005) and 10q26 (Majewski *et al*., 2003; Seddon *et al*., 2003c; Kenealy *et al*., 2004; Iyengar *et al*., 2004; Jakobsdottir *et al*., 2005).

1.5.2 AMD and the complement pathway

The earliest links between the innate immune system and AMD susceptibility came from proteomic and biochemical analyses of the drusen obtained from eye donors (Hageman *et al*., 2001; Crabb *et al*., 2002). Identification of complement components C3, C5b-9, and inhibitors of the membrane attack complex in the drusen suggested that specific immune-mediated pathways may play a significant role in the process of drusen biogenesis (Hageman *et al*., 2001; Crabb *et al*., 2002).

Complement is a central part of innate immunity, which mediates the immune response to infectious agents and links innate and adaptive immunity (Walport, 2001). The alternative pathway of the complement cascade recognises, attacks and

eliminates microbes or modified host cells, therefore it needs to be targeted to foreign surfaces and prevented from being activated on host cells and tissues (Walport, 2001). The pathway is initiated in the fluid phase by the formation of C3 by the enzyme C3 convertase, which cleaves C3 into C3a and C3b (**Figure 1.13**). The alternative pathway is non-discriminatory and the initial enzymatic components such as C3 are deposited on tissue surfaces but do not differentiate between activator and non-activator surfaces.

In order to avoid damage, host cells need to actively and continuously downregulate amplification of the complement cascade (Zipfel *et al*., 2006). There are four regulators of the alternative pathway which are expressed on the surface of host cells (CR1, MCP, DAF and CD59). Two potent regulators are also found in the fluid phase, complement factor H (CFH) (150 kDa, plasma concentrations of 110-615 μg/ml (Esparza-Gordillo *et al*., 2004)) and factor H-like protein 1 (FHL-1). The full-length human CFH is a multidomain, multifunctional glycoprotein (Ripoche *et al*., 1988) which binds C3b, accelerates the decay of the alternative pathway C3 convertase (C3bBb), acts as a cofactor for factor I-mediated proteolytic degradation of C3b and competes with factor B for binding to C3b together leading to down-regulation of the alternative complement pathway (**Figure 1.13**) (Whaley & Ruddy, 1976). FHL-1 is an alternatively spliced product of the *CFH* gene with a molecular mass of 43 kDa (Misasi *et al*., 1989).

The identification of the complement factor H (*CFH*) gene variants influencing AMD is one of the best examples of success in genome-wide association studies (GWAS) of common, complex disorders. *CFH* rs1061170 (Y402H) is a common variant that was initially identified in a genome-wide association scan for polymorphisms associated with AMD (Klein *et al*., 2005; Edwards *et al*., 2005; Haines *et al*., 2005; Hageman *et al*., 2005). These studies have shown that the *CFH* Y402H variant increases susceptibility to AMD with odds ratios (ORs) of 2.45 (one copy) and 5.57 (two copies) and the population attributable risk for AMD is estimated between 24% (Schmidt *et al*., 2006) and 54% (Despriet *et al*., 2006).

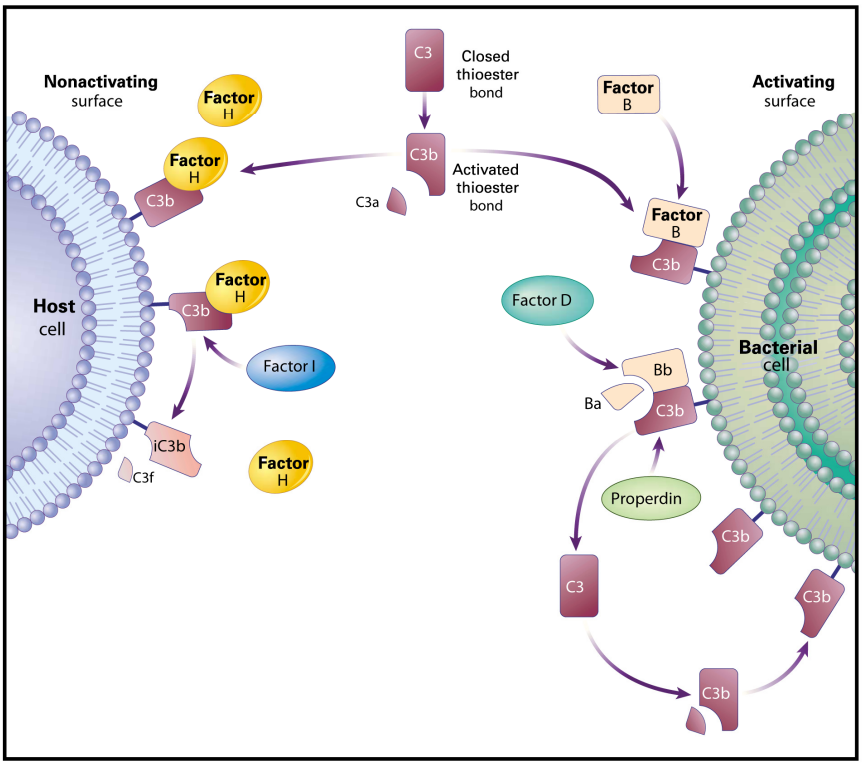


Figure 1.13 Regulation of the Cleavage of C3 by Factor H and Factor I.

The first product of the cleavage of C3 by a C3 convertase is C3b, which has an activated internal thioester bond. This bond enables C3b to bind covalently to hydroxyl groups on nearby carbohydrates and protein-acceptor groups. If the acceptor molecule is on a host cell surface, then protective regulatory mechanisms come into play. This is illustrated by the binding of factor H to C3b, which acts as a cofactor to the serine esterase factor I. Factor I cleaves the C3 into an inactive product, iC3b, releasing a small peptide, C3f. The iC3b can no longer participate in the formation of a C3 convertase enzyme. If C3b binds covalently to a bacterium, then the enzyme precursor factor B binds to the C3b. Factor B that is bound to C3b is susceptible to cleavage and activation by the enzyme factor D. This leads to the formation of the C3 convertase enzyme C3bBb, which is stabilized by the binding of properdin. This enzyme cleaves more C3, leading to the deposition of additional C3b on the bacterium. The carbohydrate environment of the surface on which the C3b is deposited determines the relative affinity of C3b for factor H or factor B. On host cell surfaces bearing polyanions such as sialic acid, factor H binds to C3b with a higher affinity than does factor B. On microbial surfaces that lack a polyanionic coating, factor B binds to C3b with a higher affinity than does factor H, leading to amplified cleavage of C3 (Diagram and text adapted from Walport, 2001).

The association between *CFH* rs1061170 (Y402H) and AMD has been replicated by many studies in different populations (Zareparsi *et al*., 2005a; Lau *et al*., 2006; Simonelli *et al*., 2006).

CFH is located on human chromosome 1q31 within the regulator of the complement activation (RCA) gene cluster which contains more than 60 genes of which 15 are complement-related genes, including *CFHR1*, *CFHR2*, *CFHR3*, *CFHR4* and *CFHR5*, to which *CFH* is closely related. CFH is a member of a group of structurally and immunologically related plasma proteins which consist of individually folding protein domains called short consensus repeats (SCRs) (Zipfel & Sherka, 1994). CFH is capable of complement regulation in both the fluid phase and on the surface of cells where it is localised by binding to cell surface polyanions such as sialic acid and heparin-like glycosaminoglycans (GAGs) (Fearon *et al*., 1978). The CFH protein also binds cellular receptors such as CD18 (DiScipio *et al*., 1998), heparin (Meri & Pangburn, 1990), the acute phase protein C-reactive protein (CRP), although not confirmed by Jarva *et al*. (1999), and the surface of certain pathogenic microorganisms such as *S. pyogenes*, *N. gonorrhoeae* and *B. afzelii* (Lindahl *et al*., 2000).

The *CFH* rs1061170 (Y402H) polymorphism occurs in the SCR7 domain and has been shown to bind heparin, using an SCR7 construct and heparin-affinity columns (Clark *et al*., 2006), and retinal pigment epithelial cells *in vitro* (Skerka *et al*., 2007). This supports the case for a causal link between the polymorphism and a disease mechanism involving insufficient complement regulation in the ageing choroid. CFH has also been found to co-localise with C3b in drusen of AMD patients (Hageman *et al*., 2005). Direct coordination of the binding between GAGs and CFH by the Y402H variant has been shown using the crystal structure of the region containing the *CFH* Y402H variant (Prosser *et al*., 2007).

The prevalence of *CFH* rs1061170 (Y402H) shows great variation between ethnic groups with a 34% frequency in the Caucasian and Somali populations and 35% in the African Americans but only 7% frequency in the Japanese and 17% in Hispanics (Lotery & Trump, 2007; Hageman *et al*., 2006). Considering the significant differences in prevalence of AMD in African Americans and Caucasians (Klein *et al*., 1999; Friedman *et al*., 2004), the similarities in frequency of the *CFH*

rs1061170 risk allele in these populations suggest that additional factors are involved in the pathogenesis of AMD.

A more detailed analysis of the *CFH* locus has also been carried out. In a study analysing 84 SNPs within and around the *CFH* gene in 544 cases of AMD and 268 age and sex-matched controls, 20 SNPs were found to be more significantly associated with AMD than *CFH* Y402H (**Figure 1.14**) (Li *et al*., 2006). Furthermore, multiple polymorphisms formed a set of 5-SNP haplotypes, using a stepwise regression analysis of all SNPs. Non-coding variants of *CFH* have been found in other studies of similar sample size to affect AMD susceptibility (Maller *et al*., 2006; Hageman *et al*., 2006).

A common deletion of two *CFH*-related genes, *CFHR3 and CFHR1* which significantly protects against AMD (independently of *CFH* Y402H) has also been discovered using 173 cases of AMD and 170 age and sex-matched control (Hughes *et al*., 2006). The deletion is present on a protective haplotype in 7.8% of AMD chromosomes and 20% of chromosomes of control individuals and the proteins encoded by these genes are absent in the plasma of homozygotes (Hughes *et al*., 2006). The *CFHR3/CFHR1* deletion polymorphisms which span approximately 86 kb, have been found to be in strong linkage disequilibrium with a protective haplotype spanning intron 11 of the *CFH* gene by Hughes *et al*. (2006). The findings of these studies have emphasised the presence of multiple susceptibility alleles in the region with non-coding *CFH* variants playing a significant role in determining disease risk.

Following the identification of *CFH* as a susceptibility locus for AMD, candidate gene studies of other components of the complement pathway have been conducted in AMD cases and controls. Variants in the complement component *C2* and complement factor B (*CFB*) were found to be associated with decreased susceptibility to AMD in multiple studies each consisting of approximately 900 AMD cases and 400 controls (ORs ranging between 0.3 and 0.5) (Gold *et al*., 2006; Maller *et al*., 2006; Spencer *et al*., 2007a).

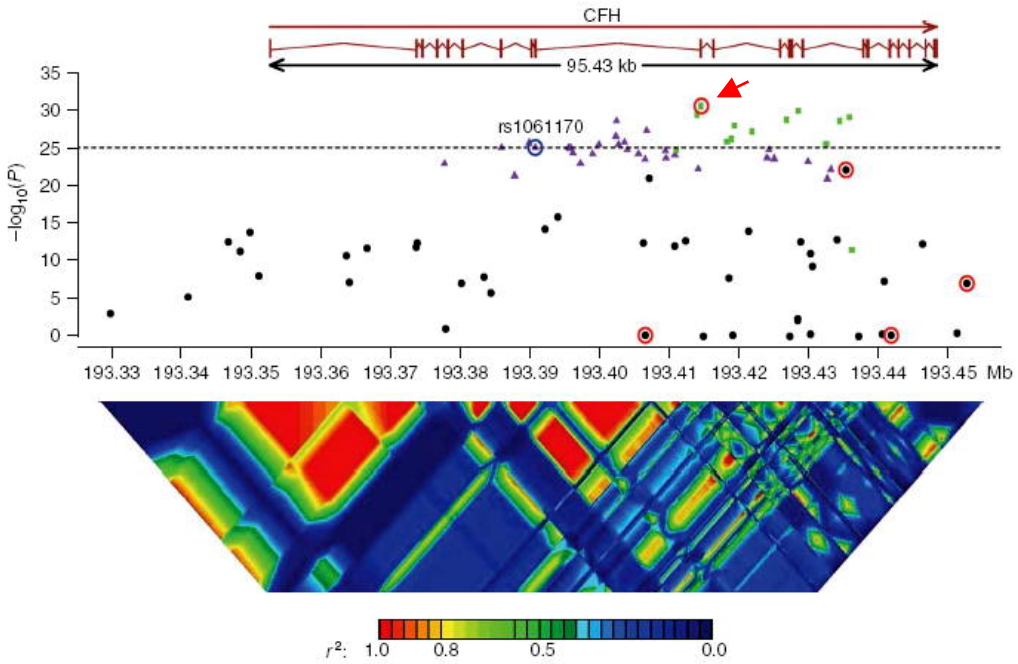


Figure 1.14 Results of the AMD association analysis of *CFH* SNPs carried out by Li *et al*., 2006.

P-values are shown for single-SNP associations, when 544 unrelated cases of AMD and 268 controls were compared. The SNP showing the strongest association with AMD (*CFH* rs2274700 (A473A), $P < 1.0E-30$) is marked by the red arrow. Twenty other SNPs showed stronger associations with AMD than the *CFH* rs1061170 (Y402H) variant. The dotted horizontal line is –log10(*P*) of the *CFH* rs1061170 (Y402H) variant ($P < 1.0E-25$) (circled in blue). Significantly associated SNPs fall into one of two LD groups ($r^2_{within\ group} > 0.80$; $r^2_{between\ groups} < 0.50$). SNPs in one of these groups are colored in green; SNPs in the other group are colored in purple; SNPs outside either group are in black. Results of this study showed that additional susceptibility alleles at the *CFH* locus may play a role in disease predisposition and *CFH* rs1061170 (Y402H) in unlikely to be the only major determinant of disease susceptibility in this region. The five SNPs (*CFH* rs1048663, rs2274700, rs412852, rs11582939 and rs1280514) selected from the stepwise haplotype association analysis are circled in red. Haplotypes containing the five SNPs were identified, three of which were shown to act independently of *CFH* rs1061170 (Y402H). Linkage disequilibrium across the *CFH* region is shown below, plotted as pairwise r^2 values. Diagram adapted from Li *et al*., 2006).

Furthermore, variation in the complement *C3* gene has recently been shown to be strongly associated with increased risk to AMD. Odds ratio of 2.6 using 603 cases and 350 controls (Yates *et al*., 2007), and OR of 3.3 using 1,238 cases and 934 controls (Maller *et al*., 2007) were reported for the homozygous allele of the common rs2230199 variant, which confers an arginine to glycine substitution at residue 80 and may influence binding between C3b/C3d and CFH (Yates *et al*., 2007). These findings have improved our understanding of age-related macular degeneration and the involvement of the complement pathway in pathogenesis of this complex disease.

1.5.3 AMD association with chromosome 10q

In addition to the *CFH* locus (1q31), significant linkage to chromosome 10q26 has been observed in individual genome-wide studies (Majewski *et al*., 2003; Seddon *et al*., 2003c; Kenealy *et al*., 2004; Iyengar *et al*., 2004; Jakobsdottir *et al*., 2005), and in a meta-analysis of AMD linkage studies (Fisher *et al*., 2005). Initial fine-mapping of this region revealed two neighbouring genes, Pleckstrin Homology-domain containing family A member 1 (*PLEKHA1*) and a hypothetical gene (*LOC387715*) (now called Age-Related Maculopathy Susceptibility 2 (*ARMS2*)) in a region of high linkage disequilibrium (LD) (Rivera *et al*., 2005; Jakobsdottir *et al*., 2005; Conley *et al*., 2006). These studies used relatively large sample sizes (N_{cases} 370-612, $N_{controls}$ 184-612) and found odd ratios of 5.7 to 8.2 attributable to the *ARMS2* rs10490924 (Ala69Ser) variant independently of *CFH* rs1061170 (Y402H).

A second polymorphism was identified in the promoter region of the HtrA serine peptidase 1 (*HTRA1*) gene in a study of "wet" cases of AMD cases and controls of Chinese origin (Dewan *et al*., 2006). The newly identified SNP, rs11200638, was associated with a 10-fold increase in AMD risk and was found to be in complete LD with the previously identified *ARMS2* rs10490924 variant. The association between SNP rs11200638 and AMD susceptibility was confirmed in a genome-wide association analysis of 581 cases of mostly wet AMD and 309 controls of Caucasian origin (Yang *et al*., 2006) which also showed that *HTRA1*, a member of a family of serine proteases, was expressed in the human retina and RPE as well as the drusen from the donor eyes of AMD patients (Yang *et al*., 2006).

More recently, analysis of 45 SNPs in the 200 kb region spanning the *PLEKHA1*, *ARMS2* and *HTRA1* genes, using 535 affected individuals and 288 controls, revealed that a single coding variant, *ARMS2* rs10490924 ($P_{association} < 10^{-30}$) can account for the association between other variants in the region and AMD, even more strongly than the SNP rs11200638 ($P_{association} < 10^{-19}$) (Kanda *et al*., 2007). Kanda and colleagues observed no promoter activity for SNP rs11200638 and no differences in *HTRA1* mRNA expression between retinas of AMD controls and controls. ARMS2 was shown to be expressed in the human retina and to localise to the mitochondrial outer membrane (Kanda *et al*., 2007).

1.5.4 Other single gene associations with AMD

Several studies have attempted to identify associations between AMD and genetic variants in candidate gene, some of which are involved in diseases showing phenotypic similarity to AMD.

Amongst candidates examined in single gene studies of AMD, *APOE* variants show the clearest association with AMD (Swaroop *et al*., 2007). Expression of the APOE protein by RPE cells has been observed (Anderson *et al*., 2001; Ishida *et al*., 2004) and APOE has been shown to localise to the drusen of AMD donor eyes (Klaver *et al*., 1998a; Li *et al*., 2006a). Furthermore, involvement of APOE in maintenance and repair of neuronal cell membranes of the peripheral and central nervous system implies a similar role in the repair of retinal damage.

Several case-control and association studies have examined the involvement of *APOE* variants in AMD susceptibility and found a lower frequency of *E4* carriers in unrelated AMD patients ($E4_{freq} \sim 0.07$) compared with age- and sex-matched controls ($E4_{freq} \sim 0.15$) in a relatively small number of AMD cases ($N \sim 100$) and controls ($N \sim 200$) (Souid *et al*., 1998; Klaver *et al*., 1998a). Although this association has been replicated in other studies (Schmidt *et al*., 2002; Baird *et al*., 2004), it seems to be in direct contrast to the proposed involvement of *APOE* in other complex disorders such as Alzheimer and coronary heart disease where *E4* is the risk allele.

The gene involved in the majority of cases of Stargardt disease, *ABCA4*, has also been tested for association with AMD, based on the phenotypic similarities

between Stargardt disease and AMD (Swaroop *et al*., 2007). A study consisting of a relatively small number of individuals showed significant associations between missense polymorphisms in *ABCA4* (D2177N and G1961E) and AMD (Allikmets *et al*., 1997 (*N* = 167 unrelated AMD cases only)). The association was confirmed in a considerably larger cohort of 1,218 AMD cases and 1,258 controls, suggesting a dominant pattern of inheritance (Allikmets, 2000). However, several studies (*N* > 200 cases) have reported negative associations between variants in the *ABCA4* gene and AMD susceptibility (Stone *et al*., 1998; Rivera *et al*., 2000; Schmidt *et al*., 2003).

Several additional reports have shown encouraging associations between AMD and the genes Cystatin C (amyloid angiopathy and cerebral hemorrhage) (*CST3*) (Zurdel *et al*., 2002), Toll-like Receptor 4 (*TLR4*) (Zareparsi *et al*., 2005b), Fibulin 5 (*FBLN5*) (Stone *et al*., 2004) and Vascular Endothelial Growth Factor (*VEGF*) (Haines *et al*., 2006). However, these studies have not been replicated, possibly due to small number of samples analysed and hence low study power.

1.5.5 Meta-analysis of established genetic associations with AMD

A recent meta-analysis of the established associations (three or more published reports) between AMD and genetic variants was carried out in which variants in the *CFH* region (*CFH* rs1061170) and in the *ARMS2* region (*ARMS2* rs10490924) demonstrated the strongest replicable association with AMD (Swaroop *et al*., 2007). Between these two loci, *ARMS2* rs10490924 was shown to confer the highest risk for AMD (OR = 2.62, $P < 10^{-100}$) (**Table 1.4**). Variants in the *APOE* (*E2*, *E3* and *E4*), *C2* (rs9332739 and rs547154) and *CFB* (rs4151667) genes also showed replicable but smaller associations with AMD across studies but significantly contribute to disease susceptibility (**Table 1.4**).

Association studies have provided valuable insights into the location of the genetic variants which influence AMD susceptibility. The advances in genotype and re-sequencing technologies have allowed genetic association studies to examine large numbers of individuals in greater detail.

Gene	Polymorphism	Total studies	Total (*N*)	Minor allele	Minor allele frequencies		Odds ratio	Meta-analysis (*P*-value)
					Cases	Controls		
CFH	rs1061170	14	10,930	C	0.565	0.365	2.00	$< 10^{-100}$
ARMS2	rs10490924	8	8,473	T	0.420	0.207	2.62	$< 10^{-100}$
C2	rs9332739	4	4,184	G	0.977	0.943	2.42	1.0×10^{-12}
C2	rs547154	4	4,162	C	0.949	0.892	2.20	6.8×10^{-10}
CFB	rs4151667	4	4,197	T	0.974	0.942	2.20	9.5×10^{-11}
APOE	-	8	4,290	*E2*	0.097	0.076	1.33	0.042
				E3	0.808	0.784	1.28	0.00024
				E4	0.095	0.152	0.60	1.7×10^{-11}

Table 1.4 A summary of conclusions from meta-analysis of established associations between AMD and genetic variants.

A meta-analysis of the established associations (three or more published reports) between AMD and genetic variants showed that polymorphisms in the *CFH* gene (*CFH* rs1061170) and in the *ARMS2* gene (*ARMS2* rs10490924) demonstrated the strongest replicable associations with AMD using combined samples of 10,930 and 8,473 individuals, respectively (Swaroop *et al*., 2007). *ARMS2* rs10490924 is shown to confer the highest risk for AMD (OR = 2.62, $P < 10^{-100}$). Variants in the *APOE* (*E2*, *E3* and *E4*), *C2* (rs9332739 and rs547154) and *CFB* (rs4151667) genes also showed replicable but smaller associations with AMD across studies. Table from Swaroop *et al*., 2007.

The development of animal models of AMD has been difficult due to the phenotypic and genetic heterogeneity of this disease in humans. Furthermore, the contribution of the environmental risk factors of AMD has not been fully assessed.

Animal studies of AMD are limited in that only primates have a macula (Hope *et al*., 1992). Transgenic pigs have some advantage because, like humans, they have a high cone to rod ratio and a slow disease process (Fauser *et al*., 2002). Rabbits have also been used but their retinal vasculature is significantly different from humans (Fauser *et al*., 2002). High cost and time commitment are disadvantages of primate, pig and rabbit animal models (Edwards & Malek, 2007). Mice, despite the lack of macula, have a higher concentration of photoreceptors centrally compared with the periphery (Karan *et al*., 2005) and show a rapid progression of disease.

Several transgenic mouse models have been developed which exhibit some of the features of human AMD such as photoreceptor degeneration and drusen formation. Examples include transgenic models of the autosomal dominant Stargardt-like macular dystrophy (expressing the mutated human *ELOVL4* gene) (Karan *et al*., 2005), and Doyne honeycomb retinal dystrophy (expressing the mutated human *EFEMP1* knock-in) (Fu *et al*., 2007). Other transgenic models including the knock-out mouse models of the Monocyte Chemoattractant Protein 1 (*Ccl2*) or the Ccl2 receptor, Chemokine CC Motif Receptor 2 (*Ccr2*) genes have been developed which show cardinal features of AMD such as accumulation of lipofuscin in and drusen beneath the RPE, photoreceptor atrophy and choroidal neovascularisation(CNV) (Ambati *et al*., 2003). The Ccr2 and Ccl2 deficient mice show that impaired macrophage recruitment allows accumulation of complement C5a and IgG which induce VEGF production by RPE, possibly mediating development of CNV (Ambati *et al*., 2003). Knock-out mouse models of other complement regulatory proteins have been developed which show the involvement of components of the immune response in AMD development. Mice deficient in the complement regulatory protein, CD59 antigen ($CD59a^{-/-}$), show increased deposition of the membrane attack complex (MAC) and develop CNV early in the disease process (Bora *et al*., 2007). Mice deficient in complement factor H ($Cfh^{-/-}$) show reduced visual acuity, increased sub-retinal deposits, thinning of the BrM and

disorganisation of the rod outer segments (Coffey *et al*., 2007). *Cfh*$^{-/-}$ knock-out mice have shown that complement factor H is critically required for the long-term functional health of the retina. An aged mouse model containing a knock-in of the human *APOE*E4* gene, combined with a high-fat diet, shows key features of dry AMD, including diffuse and focal sub-RPE deposits, thickened BrM and hyper-pigmentation of the RPE (Malek *et al*., 2005). This model combines three important AMD risk factors, including *APOE*E4*, age and a high-fat diet. Mice deficient in the Superoxide Dismutase 1 (*Sod1*$^{-/-}$) gene have features typical of human AMD (Imamura *et al*., 2006). Older animals show drusen, thickened Bruch's membrane and choroidal neovascularisation. The number of drusen increases with age and the *Sod1*$^{-/-}$ RPE cells show oxidative damage and their β-catenin mediated cellular integrity is disrupted, suggesting a role for oxidative stress in age related retinal degeneration.

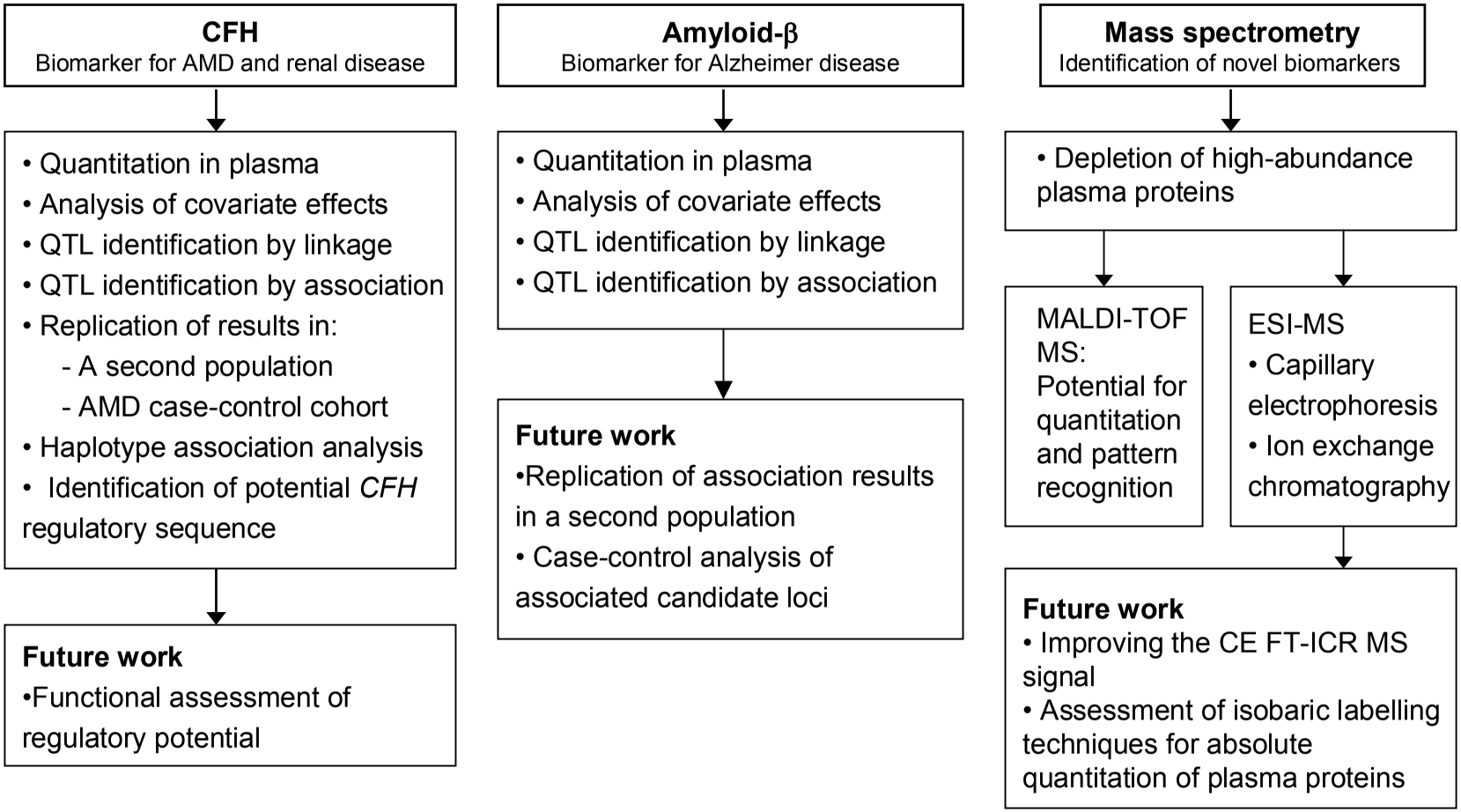
1.6 Aims

The principle aim of this project is to identify and quantitate biomarkers of disease using human plasma and to establish associations with genetic and environmental factors which influence those biomarkers (**Figure 1.15**). This is based on the assumption that intermediate phenotypes of complex diseases have a simpler genetic architecture which results in added power in quantitative trait analysis compared with binary traits (Falchi *et al*., 2004).

Figure 1.15 Project outline

Figure 1.15 Project outline

Identification and quantitation of disease biomarkers



Two groups of potential biomarkers of disease will be analysed, Aβ peptides and complement factor H (CFH) in plasma samples collected from the Vis population.

Firstly, plasma Aβ40 and Aβ42 will be used as biomarkers for Alzheimer disease, since plasma Aβ concentration is believed to reflect a global measure of Aβ toxicity in other tissues, particularly the brain (Mayeux *et al*., 1999) where Aβ42 deposition is the pathological hallmark of Alzheimer disease (Glenner & Wong, 1984; Masters *et al*., 1985). Concentrations of Aβ species in plasma will be measured for the first time in a general population, using optimised assays developed for use in EDTA anticoagulated plasma samples (Suzuki *et al*., 1994). Associations between plasma Aβ and potential covariates will be assessed. Plasma Aβ heritabilities will also be estimated using the pedigree information. Finally, genome-wide linkage and association analyses will be conducted to identify genomic loci which affect plasma Aβ concentrations in a general population.

Secondly, plasma CFH will be used as a surrogate biomarker for age-related macular degeneration (AMD) since variants in the *CFH* gene have been shown to influence susceptibility to AMD (Klein *et al*., 2005; Edwards *et al*., 2005; Haines *et al*., 2005). Association between potential covariates and plasma CFH will be assessed and the heritability of this trait will be estimated using the Vis pedigrees. Genome-wide linkage and association analyses will be conducted with the aim of identifying common genetic variants which affect plasma CFH concentration in a general population sample, which could influence susceptibility to AMD. Furthermore, genomic loci found to be associated with AMD, such as the *CFHR3/CFHR1* deletion polymorphism (Hughes *et al*., 2006) will be genotyped in the Croatian DNA samples using a novel method. Involvement of other *CFH* polymorphisms (Li *et al*., 2006) in AMD and in relation to plasma CFH concentration will be assessed in series of Scottish and Japanese AMD cases and controls after measuring plasma CFH concentrations in those samples. Replication of the CFH association results will be carried out using DNA and plasma from a Dutch isolate population as part a European Union Framework collaboration (EUROSPAN) to identify disease susceptibility variants in isolated populations.

Finally, assessment of mass spectrometric techniques will be carried out with the aim of developing methods for identification and quantitation of new disease biomarkers. The availability of large numbers of plasma samples provides a rich resource for the identification of novel biomarkers. Firstly, depletion of highly abundant plasma proteins such as Albumin, shown to interfere with the detection of lower abundance proteins (Pieper *et al*., 2003), will be carried out. Secondly, two approaches to protein analysis, "top-down" and "bottom-top", will be examined in relation to protein quantitation in plasma (Chait, 2006). If successful, large-scale protein quantitation of proteins and peptides in plasma will be used to identify novel biomarkers in population samples. Protein concentrations can then be used to estimate heritability of peaks and to unravel patterns of normal variation.

CHAPTER 2

MATERIALS AND METHODS

2.1 Introduction

In this chapter, details of experimental and analytical protocols, and the materials used are presented. The description of methods has been divided into four main sections, molecular genetic techniques, biochemical methods, proteomic methods and statistical methods used in the analysis of results. A complete list of the materials used is also given at the end.

All experimental procedures involving the use of human blood and plasma were carried out at containment level 2, according to health and safety guidelines. All experiments involving PCR were undertaken in designated areas, using sterile techniques to avoid contamination. All primers and probes were kept at stock concentrations of 100 μM at –20^{o}C and were aliquoted according to need.

2.2 Sample collection and description

Prior to the start of this project, fasting blood samples had been collected from 1,042 adult volunteers (age range between 18 and 93 years), randomly recruited from the two villages of Komiza (*N*=584) and Vis (*N*=458) situated on the Dalmatian island of Vis (**Table 2.1**). All participants had given informed consent. 1,029 unselected individuals had complete data for more than 50 disease-related quantitative traits. Of these, 8 were replicates since the same individuals had been sampled twice, giving a total population survey number of 1,021. Plasma had been separated from each blood sample and DNA extracted by colleagues in Croatia. Samples of citrate and EDTA (ethylenediaminetetraacetic acid) anticoagulated plasma, serum, and DNA from each individual were sent to MRC Human Genetics Unit in Edinburgh. Plasmas were stored at –80^{o}C upon arrival. DNA samples were diluted to a concentration of 10 ng/μl in TE (10 mM Tris HCl, 1 mM EDTA, pH 7), aliquoted into 96-well 0.8 ml storage plates (Abgene) and kept at 4^{o}C.

A Dutch population sample set, consisting of 500 DNA and EDTA anticoagulated plasma samples (**Table 2.1**) was used as a replication series for CFH analysis. These samples were from the Rucphen (Erasmus Rucphen Family (ERF)) study and consisted of unrelated individuals from an isolated village community outside Rotterdam.

	Croatian data set	Dutch data set
N	1,021	500
Mean age (SD)	56.1 (15.6)	52.9 (13.8)
% female	58.5	61.0
BMI (SD)	27.3 (4.3)	n.a.
% current smoker	22.5	n.a.
% ex-smoker	49.0	n.a.
% never smoked	28.5	n.a.

Table 2.1 Descriptive statistics of sample datasets.

Details of the Croatian and Dutch sample sets (left), and the Scottish and Japanese cohorts of unrelated cases of AMD and controls are presented. Details of the AMD grading criteria have been described in the text. N.A., not available/applicable.

		Scottish case-control data set				Japanese case-control data set	
		382 cases	201 controls			30 cases	58 controls
AMD Grade	*0*	n.a.	199	AMD Grade	*1*	n.a.	58
	1	58			*2*	2	
	2	102			*3*	1	
	3	44			*4*	3	
	4	178			*5*	24	
AMD Grade	*Normal*		201				n.a.
	ARM	160				n.a.	
	CNV/GA	222				n.a.	
Mean age (SD)		75.6 (9.5)	73.1 (8.0)			77.3 (6.4)	76.6 (5.9)
% female		62.8	62.2			16.7	60.3
% current smoker		11.7	13.3			n.a.	n.a.
% ex-smoker		38.6	45.1			n.a.	n.a.
% never smoked		49.7	41.5			n.a.	n.a.

A series of unrelated Scottish cases of AMD and controls were used for the CFH analysis (**Table 2.1**). The Scottish series comprised 382 case subjects with AMD ("dry" or "wet") and 201 age- and sex-matched control subjects (Yates *et al*., 2007). Subjects were examined by an ophthalmologist, fundus photographs were taken, and data collected regarding medical history, lifestyle, and smoking history. A total of 337 case subjects from the Lothian region were recruited from ophthalmic clinics in Edinburgh and 45 case subjects from hospitals in Dundee and Inverness, between 2004 and 2006. Age- and sex-matched control subjects were recruited from the same location and sources comprised 27 spouses and 174 subjects who had undergone cataract surgery. DNA and EDTA anticoagulated plasma samples were stored in a similar way to the Croatian samples. AMD grading was carried out by Dr. A. M. Armbrecht using the International Classification of Age-related Maculopathy and Macular Degeneration as follows (Bird *et al*., 1995); Grade 1: mild dry AMD in both eyes, grade 2: moderate dry AMD in one or both eyes, grade 3: severe dry AMD, with geographical atrophy, in one or both eyes, grade 4: severe wet AMD in one or both eyes. AMD diagnosis was performed in ophthalmic clinics in Edinburgh, Dundee and Inverness, and classification into AMD subtypes, such as choroidal neovascularisation or geographic atrophy, was carried out. This series was part of a larger association study of age-related macular degeneration (Yates *et al*., 2007). All individuals gave informed consent to the study.

A series of 30 unrelated Japanese cases of AMD and 58 age- and sex-matched controls recruited by Dr. T. Iwata from the National Institute of Sensory Organs were also obtained for the population-specific analysis of *CFH* gene variants (**Table 2.1**). The grading system used for the clinical diagnosis and classification of AMD amongst cases was based on the Clinical Age-Related Maculopathy Staging (CARMS) as follows (Seddon *et al*., 2006b); grade 1: no drusen or < 10 small drusen, grade 2: approximately ≥ 10 small drusen, RPE hypo/hyperpigmentation, grade 3: approximately ≥ 15 intermediate drusen, grade 4: geographic atrophy, grade 5: choroidal neovascularisation with RPE detachment. All individuals gave informed consent to the study.

2.3 Molecular genetic techniques

2.3.1 *APOE* genotyping

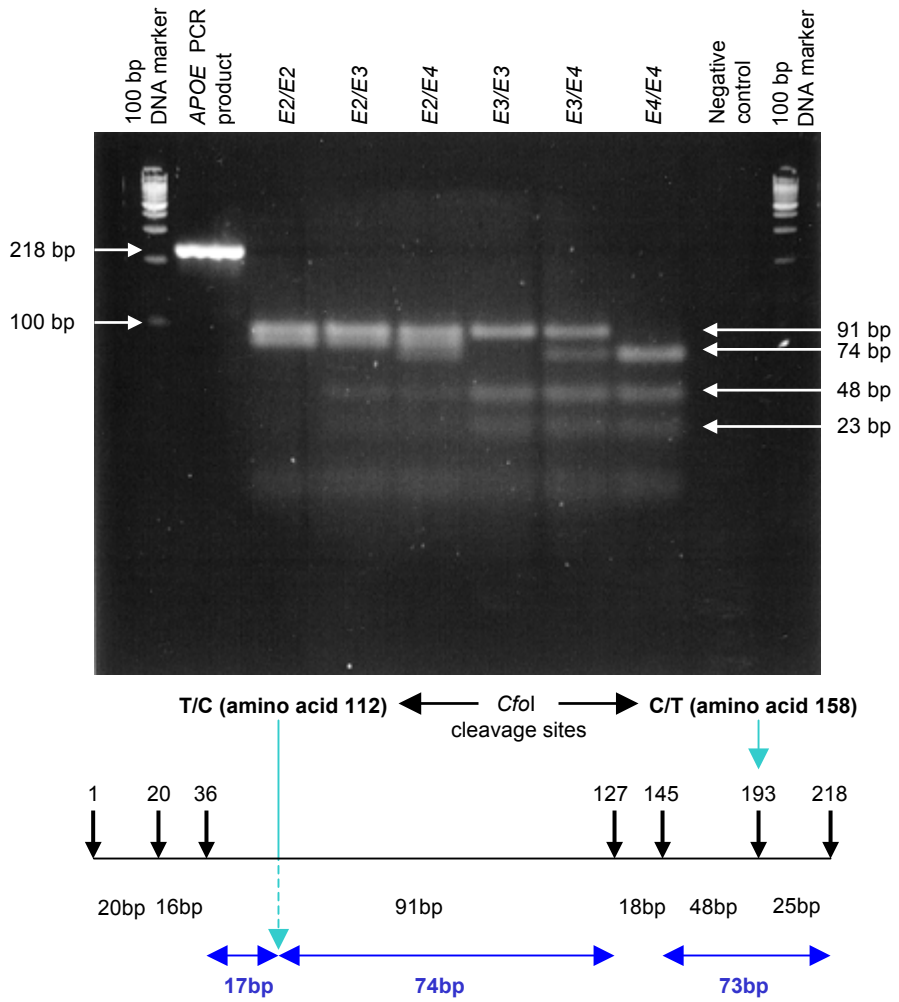
APOE genotyping was carried out in the Vis population samples by polymerase chain reaction (PCR) amplification of DNA samples using a pair of oligonucleotide primers (Invitrogen) designed to amplify a 218 bp region spanning the two *APOE* variant sites112 (T > C) and 158 (C > T) (*APOE* forward primer: 5′-TCCAAGGA-GCTGCAGGCGGCGCA and *APOE* reverse primer: 5′-ACAGAATTCGCCCCGG-CCTGGTACACTGCCA) as shown in **Figure 2.1**. The composition of the PCR using a 2x Thermo-Start® PCR Master Mix (ABgene) was as follows:

- 1.25 units Thermo-Start® DNA Polymerase
- 1x Thermo-Start® reaction buffer
- 2.5 mM $MgCl_2$
- 0.2 mM each of dATP, dCTP, dGTP and dTTP

	Volume/Reaction (µl)	Final Concentration
Thermo-Start PCR Master Mix (2X)	16.5	Stock diluted 1:2
Dimethyl Sulfoxide (2X) Sigma®	2.5	Stock diluted 1:2
Forward Primer (2 µM)	2.5	0.2 µM
Reverse Primer (2 µM)	2.5	0.2 µM
DNA template (10 µg/ml)	1.0	-

Table 2.2 Composition of the PCR master mix used in the amplification of APOE DNA fragment.

PCR was carried out in 0.2 ml semi-skirted 96-well plates (Thermo-Fast® 96, AB gene) in a total volume of 25 µl per well. 2 µl of Orange-G® Loading Buffer was added to each PCR sample. The final mix was loaded in a 2% agarose (Invitrogen) minigel and run at 100 V for approximately 45 minutes. A 100 bp DNA size marker (Promega) was also run on each gel. The bands were visualised using a UV transilluminator (Syngene).



	Nucleotide (Amino acid residue)	
	112	158
*APOE*E2*	T (Cys)	T (Cys)
*APOE*E3*	T (Cys)	C (Arg)
*APOE*E4*	C (Arg)	C (Arg)

APOE ^{112} T -> C (rs429358)
APOE ^{158} C -> T (rs7412)

Figure 2.1 Illustration of the APOE genotyping.

DNA of each of the six possible *APOE* genotypes, together with a full length 218bp amplified *APOE* sequence containing the two variants digested with *Cfo*I captured on a 4% MicroSieve agarose gel, stained with ethidium bromide is shown. Diagram and table below illustrate the *APOE* digestion at two cleavage sites (light blue arrows). Note that *Cfo*I cleaves the DNA if a Cytosine base is present at the site. Black vertical arrows show other *Cfo*I cleavage sites (*APOE*E2*: TT, *APOE*E3*: TC, *APOE*E4*: CC).

2.3.1.1 *APOE* genotyping by RFLP

Genotyping was carried out by restriction fragment length polymorphism (RFLP) analysis. This consisted of overnight digestion of the *APOE* PCR product at 37°C, using the restriction endonuclease enzyme *Cfo*I (Roche), which is an isoschizomer of *Hha*I that cleaves either within residue 112 or 158 (**Figure 2.1**). *Cfo*I recognises the sequence GCG/C and generates fragments with 3′ cohesive termini. Digestion was carried out in 15 µl reactions (**Table 2.3**). Five microlitres of the digestion mix was added to 10 µl of PCR product in each well of a 0.2 ml semi-skirted Thermo-Fast® 96 plate (ABgene). Separation of digested fragments was carried out by agarose gel electrophoresis. Four percent MicroSieve Low Melt agarose (Flowgen) gels were prepared in 0.5x TBE Buffer containing 0.002% of 10 mg/ml ethidium bromide (BDH®). Electrophoretic separation was carried out for approximately 3 hours at 60 volts.

	Volume/reaction (µl)
10x SuRE/Cut buffer L	1.50
Cfo I enzyme (10u/µl)	0.15
Gibco™ Water	3.35
PCR product	10.00

Table 2.3 Composition of the CfoI digest for *APOE* genotyping

Genotyping analysis was carried out based on the number and size of the bands observed on the gel (**Figure 2.1**). The combination of bands expected for each genotype (Wenham *et al*., 1991) was obtained using the programme Restriction-Mapper (http://www.restrictionmapper.org) (**Table 2.4**).

APOE Genotype	Fragment sizes observed after RFLP analysis (bp)
E2/E4	91, 74, 73, 48, 25, 20, 18, 17, 16
E2/E3	91, 73, 48, 25, 20, 18, 16
E2/E2	91, 73, 20, 18, 16
E3/E4	91, 74, 48, 25, 20, 18, 16
E3/E3	91, 48, 25, 20, 18, 16
E4/E4	74, 48, 25, 20, 18, 17, 16

Table 2.4 Expected fragment sizes of amplified *APOE* after digestion with CfoI.

2.3.1.2 Sequencing of *APOE*E2* carriers

Due to the uncertainty associated with deducing the *APOE*E2* genotypes by RFLP, putative carriers of the *E2* allele (i.e. *E2*/*E2*, *E2*/*E3* and *E2*/*E4* genotypes) were sequenced in both directions using BigDye™ Terminator Cycle Sequencing version 3 (Applied Biosystems®) (**Table 2.5**). The aim was firstly to confirm the presence of the *E2* allele in the *APOE*E2* carriers, and secondly to differentiate between the *E2*/*E2* and *E2*/*E3* genotypes.

	Volume (µl/reaction)	Concentration
Gibco™ H_2O	6.0	-
BigDye	2.0	-
Primer	1.0	40 nM

Table 2.5 Composition of the BigDye sequencing reaction mix.

Prior to sequencing reaction, 5 µl of each *APOE* PCR product was purified using 1 µl of 1 unit/µl Shrimp Alkaline Phosphatase (SAP, USB) and 0.5 µl of 10 unit/µl exonuclease I (USB). The phosphatase reaction required incubation with SAP for 15 minutes at 37°C followed by 15 minutes at 80°C in a thermal cycler (MJ Research). The sequencing reaction was then carried out using 2 µl of the previously purified DNA template, in a 11 µl reaction volume (**Table 2.5**), by amplification using 25 cycles of 30 seconds at 96°C, 15 seconds at 50°C and 4 minutes at 60°C.

The sequencing reaction was followed by DNA precipitation, by adding 0.1 v/v NaOAc and 2.5 v/v of 95% ethanol (3M, pH 4.0). This was followed by centrifugation (Multifuge 3) at 2000 g for 30 minutes at 20°C. The pellet was washed once with 150 µl of 70% ethanol, dried and stored at –20°C. After automated sequencing on an ABI PRISM® 3100 (HGU core facility), the generated DNA sequences were analysed using the software Sequencher® v3.0 (Gene Code).

2.3.1.3 Experimental controls

Experimental controls were included throughout the genotyping reactions. Negative controls included blank wells for PCR and digestion experiments. Positive controls consisted of samples genotyped confidently as *E3/E3* and *E4/E4*. These were also included in the sequencing experiments.

2.3.1.4 Genotyping replications

A set of 95 random DNA samples from each of the two villages of Komiza and Vis was used to confirm the respective genotype of each individual by genotyping a second time. Samples were selected based on their identification numbers using a random number generator with Microsoft Excel® software.

2.3.2 Genotyping of STR markers

Genome-wide linkage analysis was carried out using 810 short tandem repeat (STR) markers from the ABI Prism® Linkage Mapping Set HD5 (Applied Biosystems). Genotyping of markers in the Vis sample set was performed by Susan Campbell and Joanne Morgan using the Gene Mapper® Software version 3.7 (Applied Biosystems). Microsatellite markers were spaced, on average, 3-5 cM apart. Genotyping errors were detected by Dr. C. Hayward using the software PedCheck which scans genotype data for inconsistencies in Mendelian inheritance. Genotype data was then prepared for linkage analysis by Dr. C. Hayward.

2.3.3 Genotyping of single nucleotide polymorphisms

Three methods were employed for the genotyping of single nucleotide polymorphisms (SNPs). Genotyping of 317,503 SNPs was carried out using the Human Hap300® Array (Illumina). SNPs which had not been included in the Illumina platform were genotyped by a 5′ nuclease assay using TaqMan® probes (Applied Biosystems), or by direct sequencing.

2.3.3.1 Whole genome SNP genotyping

317,503 SNPs were genotyped as part of the Human Hap300® (Illumina, San Diego, USA) genotyping platform in 986 Croatian individuals. The array included all tagging SNPs from the HapMap Phase I SNPs, 8,000 non-synonymous SNPs and 1,500 tagging SNPs within the human major histocompatibility complex (MHC). Genotypes were scored using the Bead Studio® software v.3 (Illumina). A total of 8,984 SNPs had less than 90% genotyping call rates and 379 had a minor allele frequency < 1% and were removed, leaving 308,140 SNPs for the follow-up genome-wide association analysis.

2.3.3.2 SNP genotyping by TaqMan® 5′nuclease assay

Five-prime nuclease assays using TaqMan® probes (**Figure 2.2-A**) were used to genotype SNPs in the *CFH* genomic region in four sample sets of Croatian, Dutch, Scottish and Japanese origin. Assays were either custom, or pre-designed, involving a probe mix containing two minor groove binder (MGB) probes, each labelled with a different fluorescence reporter dye (FAM, VIC). Custom-designed assays involved electronically providing around 750 bp of DNA containing the SNP of interest, to the supplier (Applied Biosystems) who designed and manufactured the primers and probes.

A list of SNPs and the population series in which they were genotyped is provided below:

- Croatian data set: rs1061170 (Y402H), rs1048663, rs412852 and rs1066420
- Dutch data set: rs1061170 (Y402H), rs6677604, rs1329428, rs7517126, rs4086175, rs3766404, rs1048663, rs412852, rs1066420 and rs11582939
- Scottish AMD data set: rs1061170 (Y402H), rs3766404, rs1048663, rs412852, rs1066420 and rs11582939
- Japanese AMD data set: rs1061170 (Y402H), rs6677604, rs1329428, rs7517126, rs4086175, rs3766404, rs1048663, rs412852, rs1066420 and rs11582939

SNP genotyping in the Japanese DNA samples was carried out by Dr. T. Iwata in Japan due to Japanese regulations on export of DNA samples. Genotyping of SNPs in the Dutch sample set was carried out by Susan Campbell.

Each reaction contained 0.25 μl of 20x TaqMan® probe mix, diluted 1 in 2 using TE (10 mM Tris HCl, 1 mM EDTA, pH 7.0) from a 40x stock (Applied Biosystems), 2.5 μl of 2x TaqMan® Universal PCR Master Mix (ABI), and 2.25 μl of Gibco™ dH_2O (Invitrogen). Optical reaction plates (384-well, Applied Biosystems) were coated with 10 ng of DNA (10 ng/μl) and left to dry overnight. The reaction was carried out on a thermal cycler (MJ Research) and consisted of 40 cycles of 10 minutes at 95°C, 15 seconds at 92°C and 1 minute at 60°C.

After completion of the PCR, fluorescence was read using an ABI-HT7900 SDS instrument (Applied Biosystems). Results of the fluorescence-labelled allelic discrimination were consequently exported and analysed further using Microsoft Excel®. The presence of FAM or VIC dye fluorescence alone indicated homozygosity for the FAM or VIC- specific allele. An increase in both signals indicated heterozygosity.

2.3.3.3 SNP genotyping by direct sequencing

Genotyping of *CFH* rs2274700 was carried out by DNA sequencing. This SNP was genotyped in the Croatian, Dutch and Scottish AMD series. Genotyping was also carried out in the Japanese AMD series in Japan. The reason for this method of SNP analysis was the presence of other polymorphisms in exon 10 of the *CFH* gene and the advantage of genotyping both SNPs.

Forward (5′ TGTCTTTGGCAACTCTGAGC) and reverse (5′-CAGCCCCCACAAAAAGACTA) primers were designed to amplify a 436 bp region including exon 10 of the *CFH* gene, using the web resource Primer 3 (Rozen & Skaletsky, 2000). PCR was carried out as shown in **Table 2.4**, using 1 μM (final concentration) of each primer at 60°C annealing temperature. DNA purification and sequencing (ABI PRISM® 3100) was carried out by the MRC HGU core facility. The generated DNA sequences were analysed using the software Sequencher® v3.0 (Gene Code).

2.3.4 Genotyping of the *CFHR3/CFHR1* deletion

The deletion of the two *CFH*-related genes, *CFHR3* and *CFHR1*, believed to be involved in protection from AMD (Hughes *et al*., 2006) was genotyped using a quantitative PCR (qPCR) assay which was developed in-house.

2.3.4.1 Theory of quantitative polymerase chain reaction

There is an exponential relationship between the concentration of DNA template material and concentration of PCR amplified product at any given cycle. qPCR is a probe-based technique involving a reporter fluorescent and a quencher dye. In the presence of the target sequence the probe anneals downstream of the forward primer and is cleaved by the 5′ exonuclease activity of *Taq* DNA polymerase during the extension phase. At each cycle, cleavage of the probe results in the separation of reporter dye from quencher which is then detected (**Figure 2.2-A**). The higher the starting copy number of the DNA template, the sooner a significant increase in fluorescence is detected above a baseline. The parameter C_t is defined as the fractional cycle number at which the fluorescence passes the fixed threshold (**Figure 2.2-A**). The threshold is set so it crosses all amplification curves where the slopes of the curves are at their highest and curves are almost linear.

There is an inverse linear relationship between the C_t value and the $\log_{10}$ of the input DNA concentration. The copy number calculation assumes a log phase amplification and a C_t value is usually taken to be about ten times the standard deviation of the baseline. Quantitation of the sample is performed relative to a control, β-globin (*HBB*) gene amplification which is valid provided that both show similar slopes during the amplification process, indicating similar amplification efficiencies. A two-fold difference in DNA copy number would result in one unit difference in the C_t value as follows,

$$\mathbf{R_c = 2^{\Delta Ct}}$$

where R_c is the relative copy number of DNA and ΔC_t is the difference in C_t of the control and test samples (Higuchi *et al*., 1993).

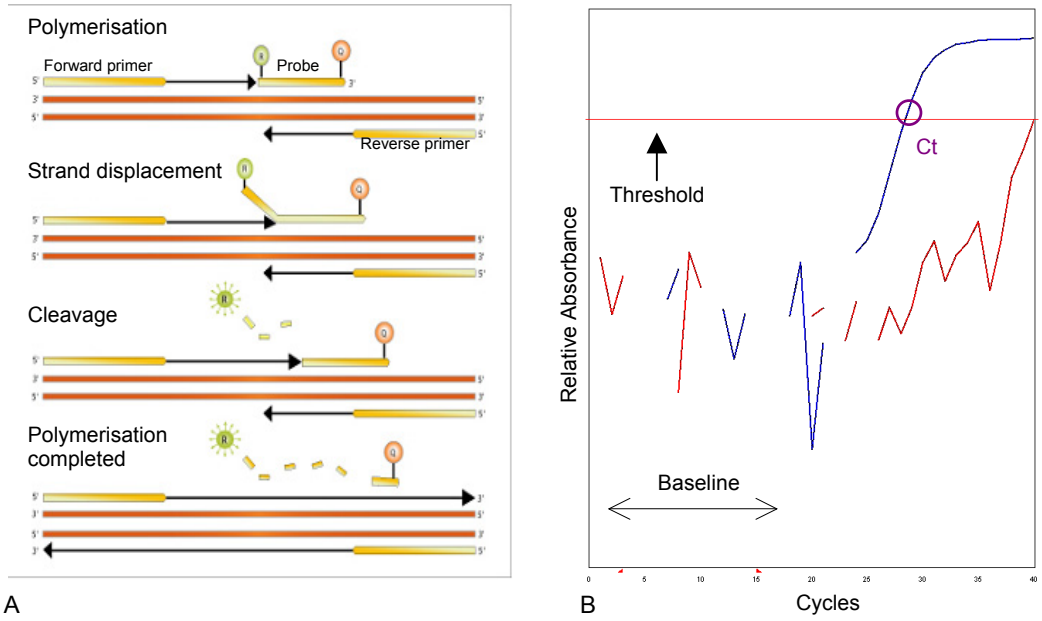


Figure 2.2 Diagrammatic presentation of the quantitative PCR methodology.

A) Stages of a qPCR reaction involving a fluorescent-labelled probe.
B) An example of the outcome showing a sample (black line) and a blank control (red line).

2.3.4.2 Sequence analysis at the *CFH* locus

Stretches of DNA which did not fall into regions of genomic duplication were identified using the sequence alignment tool BLAST2 (Tatiana *et al.*, 1999), as shown in **Figure 2.3**. BLAST2 sequence analysis revealed a high degree of sequence homology between *CFHR4* and *CFHR3*, shown in black, blue and green diagonal boxes (**Figure 2.3**). Two regions of 3,764 bp and 828 bp were found to contain non-homologous sequence. These two regions are marked by yellow and purple rectangles in **Figure 2.3**. The software Primer Express® (Applied Biosystems) was used for designing optimal primers and probes in the 828 bp region. The betaglobin gene, *HBB* (11p15) was used as an endogenous control to provide relative quantitation of *CFHR3* allele dosage.

Genotyping of the *CFHR3*/*CFHR1* deletion polymorphism was carried out in two stages. Firstly, all individuals homozygous for the deletion were identified using

a multiplex PCR. Secondly, a high throughput real-time qPCR was developed and optimised for the genotyping of the deletion aimed at distinguishing between heterozygous carriers of the deletion and homozygous non-carriers. The deletion was genotyped in the Croatian, Dutch and Scottish samples.

2.3.4.3 Multiplex PCR of *CFHR3/CFHR1* deletion

Two pairs of primers were designed using the web resource Primer 3 (Rozen & Skaletsky, 2000). One pair was designed to amplify a 219 bp product from intron 3 of the *CFHR3* gene, located in the deleted region (forward primer, 5′ TGTTTTGCC-AACGGACCTAT; reverse primer, 5′ TTCTTGGTGCAAGATGACGA). A second pair of primers was designed to amplify a 280 bp product from the *HBB* gene (11p15), for use as an endogenous control in the multiplex PCR (forward primer, 5′ CAACTTCATCCACGTTCACC; reverse primer, 5′ GAAGAGCCAAGGACAG-GTAC).

PCR amplification was carried out in 96-well semi-skirted plates (Abgene). Ten nanograms of DNA (10 ng/μl) was used as template in 13 μl reactions containing 7 μl of 2x Multiplex PCR Master Mix (Qiagen) and 1.25 μl of each of the four primers (10 μM concentration). Reaction conditions for the multiplex PCR amplification were as described before using an annealing temperature of 62^{o}C. After completion of PCR, 2 μl of Orange G® Loading Buffer was added to each sample. Separation of PCR products was carried out by agarose gel electrophoresis. Agarose gels (3%) were prepared using Hi Pure® agarose (Biogene) in 0.5x TBE buffer containing 0.002% of 10 mg/ml ethidium bromide (BDH). Electrophoretic separation was carried out for approximately 2 hours at 100 V. A 100 bp DNA size marker (Promega) was also run on each gel. The bands were visualised using a UV transilluminator (Syngene).

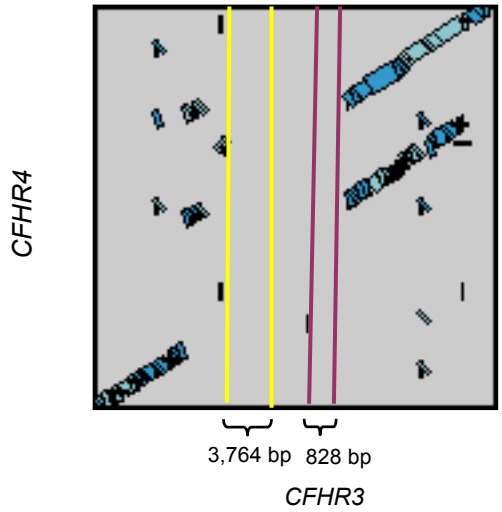


Figure 2.3 BLAST2 sequence analysis of the *CFHR4* and *CFHR3* genes.

Result of sequence alignment of the genomic sequences of both *CFHR4* and *CFHR3* showing stretches of DNA homology. The two candidate regions of non-duplicated DNA used to design primers and probes are marked by yellow and purple lines. Black, blue and green segments show varying levels of homology between the two sequences.

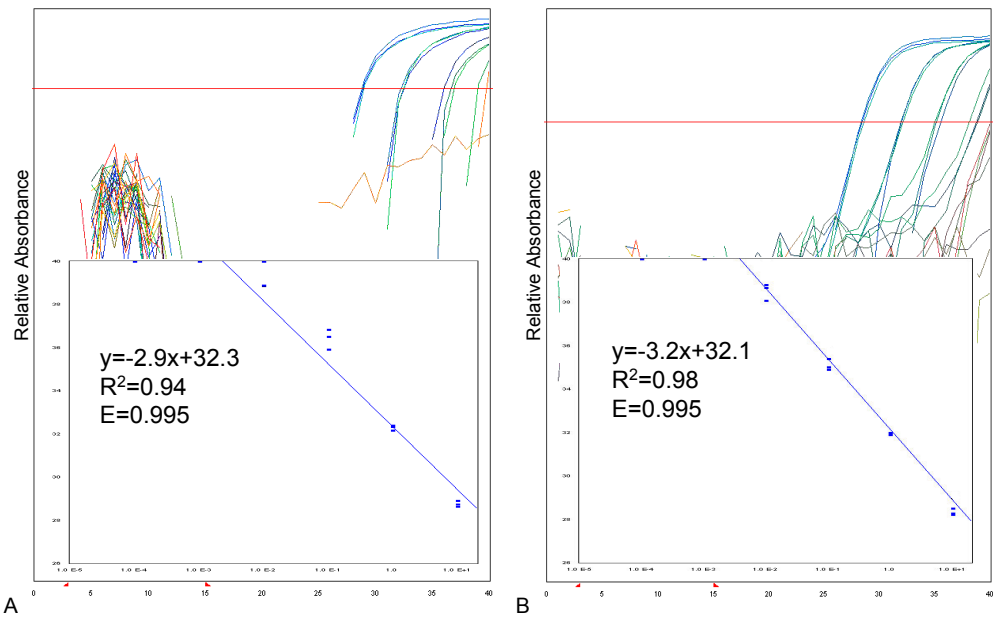


Figure 2.4 Optimisation results of probe, primer and DNA template concentrations.

Amplification and standard plots for the FAM **(A)** probe for the *CFHR3* gene and VIC-labelled **(B)** probe for the *HBB* gene using ten-fold dilutions of sample DNA template. All sample dilutions were carried out in triplicates. The efficiency of each reaction was calculated as E. Equations for each standard curve and the squared correlation coefficients (R^2) are also shown.

2.3.4.4 Quantitative PCR of *CFHR3/CFHR1* deletion

2.3.4.4.1 Designing the qPCR primers and probes

Two sets of primers and minor groove binder (MGB) probes (Applied Biosystems) were designed using the software, Primer Express® v.2.0 (Applied Biosystems). Each set consisted of a pair of primers and a 3′-fluorescent-tagged probe (**Table 2.6**). One pair of primers and a FAM™-labelled probe were designed inside intron 3 of *CFHR3*. Another pair of primers and a VIC™-labelled probe was designed inside the *HBB* gene and used as an endogenous control.

Name	Primer Sequence (5′ to 3′)	5′-Fluorescent dye
CFHR3-F	TGGGCATTAGTCAAGAATACAGTAAAA	-
CFHR3-R	ATTAATGCCGCTTCAATATGACTTT	-
CFHR3-Probe	AATTAGAACACAATACTTGTTGGC	6-FAM™
Bglobin-F	GGGCAGAGCCATCTATTGCTT	-
Bglobin-R	TGGTGTCTGTTTGAGGTTGCTAGT	-
Bglobin-Probe	TTGCTTCTGACACAACTG	6-VIC™

Table 2.6 Details of primer and MGB probes used in the quantitative PCR of *CFHR3/CFHR1* deletion

2.3.4.5 Optimisation of primer and probe concentrations

Optimal concentrations of primers and probes in both *CFHR3* and *HBB* genes were determined by analysing their standard curves (**Figure 2.4**). A duplex reaction was used in which beta globin and *CFHR3* were simultaneously amplified in the same reaction tube.

The qPCR was optimised using varying concentrations of primers and probes to suit a multiplex set-up with minimum inter-experimental variation. The optimisation procedure consisted of using 10-fold dilutions of a randomly selected DNA sample and detecting the accuracy with which quantitation would take place against a standard curve. Optimal fluorescence was observed at 10 ng/ml of DNA template with the least amount of variation between sample replicates (**Figure 2.4**). Furthermore, efficiency of each reaction was determined using the 'Full Efficiency Theory' (Applied Biosystems) as follows:

$$E=10^{(-1/slope)}-1$$

E is the efficiency of the reaction and slope refers to the slope of the standard curve. Therefore the closer the slope of the standard curve is to –3.32 (E=100%) the higher the efficiency of the reaction. The efficiency of amplification reactions at the selected template concentrations was above 99.5% for both *CFHR3* and *HBB*.

2.3.4.5.1 Experimental approach

Each qPCR was carried out in triplicate using 384-well optical reaction plates (Applied Biosystems). 5 µl reactions contained 1 µl of DNA (10 ng/ml) used as template, 2.5 µl of 2x TaqMan® Universal PCR Master Mix (Applied Biosystems), 0.2 µl of each primer at 10 µM (**Table 2.6**), 0.2 µl of each probe at 1 µM (**Table 2.6**), and 0.3 µl of Gibco™ dH_2O (Invitrogen). Care was taken to minimise experimental variation caused by pipetting errors and inconsistencies. Reactions were carried out real-time, using the absolute quantitation (standard curve) setting on an ABI-HT7900 SDS instrument (Applied Biosystems). Conditions used in the qPCR are as follows, 2 minutes at 50°C, 10 minutes at 95°C, and 40 cycles of 15 seconds at 95°C and 1 minute at 60°C. After completion of PCR, fluorescence was read using the software SDS (Applied Biosystems) and the resulting C_t values were exported to and analysed using Microsoft Excel®.

2.4 Biochemical techniques

2.4.1 Commercial assay of amyloid-β in plasma

Measurements of Aβ40 and Aβ42 concentrations were initially carried out in 88 EDTA plasma samples from the island of Vis, using sandwich ELISA kits supplied by Biosource UK, following the manufacturer's instructions.

The Aβ40 and Aβ42 standards used in the experiment consisted of lyophilised synthetic peptide (Biosource UK) which was reconstituted before use, in Standard Reconstitution Buffer. A protease inhibitor cocktail containing 4-(2-

Aminoethyl) benzenesulfonyl fluoride (AEBSF) (Sigma) was used at a final concentration of 1mM. Initially plasma samples (diluted 1:5) were tested in duplicate for assays of Aβ40 and Aβ42 peptides. Colour development was achieved using 100 µl of Stabilised Chromogen (Biosource UK) per reaction well and reactions were terminated using a Stop Solution (Biosource UK) after 30 minutes. Photometric detection was carried out at 450 nm using the SkanIt plate reader (Thermo Electron) within 20 minutes of stopping the reactions. Results were analysed using the SkanIt software version 2.1.84 (Thermo Electron) and Microsoft Excel®.

Analysis consisted of subtracting the blank from the rest of the A_{450} optical density/absorbance readings followed by construction of a standard curve for each plate, which was used to determine the concentration of Aβ40 or Aβ42 in each sample. Inter and intra-experimental variation was also calculated using the coefficient of variation.

2.4.2 Optimised assay of amyloid-β in plasma

A 'sandwich' ELISA, developed by Dr. N. Suzuki and optimised in the laboratory of Professor S. G. Younkin (Mayo Clinic, Jacksonville, USA) was used for *in vitro* quantitative determination of human plasma Aβ40 and Aβ42 concentrations. Analysis was carried out in triplicate using EDTA plasma samples collected from the isolated island of Vis. Details of all buffer compositions are provided in section 2.6.

The experimental procedure was carried as follows; centre wells of 96-well microtitre plates (Greiner Bio-one, US) were coated with 100 µl of 5 µg/ml monoclonal capture antibodies BAN50 (Aβ40) and BNT77 (Aβ42), diluted in Coating Buffer. Plates were incubated overnight at 4°C and blocked with 300 µl of blocking solution, Block Ace® (Serotec Ltd.) on the following day after discarding the unbound capture antibodies. Incubation with the blocking agent was carried out overnight at 4°C. All wells were then washed twice with 1x PBS buffer (pH 7.4). Fifty microlitres and 116.7 µl of Buffer EC were added to 'standard' and sample wells respectively, in order to prevent drying of the wells. Synthetic Aβ40 and Aβ42 standards at a stock concentration of 5000 fmol/ml were diluted using Buffer EC to a

starting concentration of 800 fmol/ml. Four serial dilutions of 1 in 4 were prepared for plasma samples and a blank control (Buffer EC). One hundred microlitres of each diluted standard was added to relevant wells in duplicate. Plasma samples which had been thawed and mixed by inverting, were diluted 1 in 3 by adding 33.3 µl of crude plasma to each well containing 116.7 µl Buffer EC in duplicate. Four plasma samples were used as internal controls on each plate. Plates were covered with SealPlate® (Applied Biosystems) and incubated overnight at 4°C. After incubation, plates were washed twice using PBS. One- hundred microlitres of horseradish peroxidase (HRP)-conjugated detection antibodies, BA27 (Aβ40) and BC05 (Aβ42), diluted 1 in 2000 using Buffer C was added to each centre well only. Plates were then incubated at room temperature for 4 hours before being washed twice with PBS and a third time using PBS-Tween® (Sigma). Colour development was achieved by incubating samples and standards with 100 µl of the Developing Solution prepared by mixing equal portions of TMB (3,3′,5,5′-tetramethylbenzidine) peroxidase substrate (KPL, Inc., US) and peroxidase Solution B (KPL, Inc., US). Plates were incubated for 30 minutes (Aβ40) and 60 minutes (Aβ42) at room temperature. Photometric detection was carried out at 450 nm using the Softmax® Pro plate reader (Molecular Devices) within 20 minutes of stopping the reactions with 100 µl of 1M phosphoric acid. Results were analysed using the Softmax® Pro software (Molecular Devices) and Microsoft Excel®.

2.4.3 Assay of complement factor H in plasma

A ‘sandwich’ ELISA was developed and optimised to perform quantitation of human complement factor H (CFH) in plasma. Measurements were carried out in EDTA plasma samples from Vis, Rucphen, Scotland and Japan (series described in **Table 2.1**). Five plasma samples collected locally were used as internal controls throughout the experiments and for normalisation of measurements. CFH quantitation in the Dutch sample set was carried out by Joanne Morgan.

2.4.3.1 Plasma extraction from blood

Blood samples were collected from local volunteers for use as internal controls in the CFH assay. Blood was taken using EDTA-coated tubes in 5-10 ml volumes. Tubes containing the extracted blood were left at room temperature for 30 minutes prior to centrifugation at 2,500 *g* for 15 minutes. Plasma was then collected and aliquoted into 20 µl fractions and stored at –80^{o}C.

2.4.3.2 CFH assay optimisation

Dilutions of OX23, a mouse monoclonal anti-CFH detection antibody (donated by Dr. R. B. Sim, University of Oxford) and the HRP-conjugated goat anti-mouse antibody (Biosource UK) were optimised for CFH assay as follows. Fifty microlitres of 1 µg/ml of purified full-length human CFH (Dr. R. B. Sim), diluted in the Coating Buffer from a 1.5 mg/ml stock, was used to coat the wells of a 96-well microtitre plate (Greiner Bio-one) in duplicate. CFH assay also included a blank control. After incubation at room temperature for 2 hours, plates were washed 4 times using the Wash Buffer. Unoccupied sites were blocked with 100 µl of 1% lyophilised bovine serum albumin (98% purity, Sigma) for 2 hours at room temperature with shaking (Grant-Bio). This was followed by 4 washes, after which plates were stored at 4^{o}C over night.

Serial dilutions of 1:1,000, 1:2,000 and 1:4,000, of the 340 µg/ml OX23 detection antibody were made using the Sample Diluent. Fifty microlitres of each dilution was added to each well and incubated at room temperature for 3 hours with shaking. Plates were washed 4 times using the Wash Buffer. Fifty microlitres of HRP-conjugated antibody was serially diluted 1:1,000, 1:2,000, 1:4,000 and 1:8,000 from the 580 µg/ml stock using the Sample Diluent. These were then added to each well as determined previously by a checker-board titration. Samples were incubated at room temperature for 30 minutes with shaking. After 4 washes, colour development was achieved by adding 100 µl of stabilised TMB Chromogen (Biosource UK) and incubation for 1 hour at room temperature with shaking. One hundred microlitres of the Stop Solution was used to terminate the reaction and absorbance was read at 450 nm using the SkanIt plate reader (Thermo Electron). The

titration curve obtained using the SkanIt software version 2.1.84 (Thermo Electron) was analysed using Microsoft Excel® and the best combination of the detection and secondary antibody dilutions were selected for further experiments.

Optimal dilution of the 9.6 mg/ml sheep polyclonal anti-CFH capture antibody (Abcam) was established by coating the wells of a microtitre plate (Greiner Bio-One) with different dilutions of the capture antibody in the Coating Buffer. Fifty microlitres of the diluted antibody was coated onto each well in duplicate. After incubation at room temperature for 2 hours, the plate was washed 4 times in the Wash Buffer and unoccupied binding sites were blocked with the Blocking Solution as before. The plate was washed 4 times and stored at 4°C overnight. Serial dilutions of the CFH standard were prepared using the Sample Diluent starting at a concentration of 300 ng/ml. A blank control was also included. Fifty microlitres of each standard was added to wells coated with the capture antibody and samples were incubated at room temperature for 2 hours using the plate shaker. After 4 washes, 50 µl of the detection antibody, at a concentration of 85 ng/ml, was added to each well and incubated at room temperature for 3 hours using the plate shaker. Fifty microlitres of the HRP-conjugated (Biosource, UK) antibody, at a concentration of 145 ng/ml, was added to each well after washing the plate 4 times. Incubation was carried out for 30 minutes at room temperature, with shaking. After washing the plate 4 times, colour development was achieved by adding 100 µl of TMB Chromogen (Biosource UK) and incubation for 1 hour at room temperature with shaking. The reaction was stopped using 100 µl of Stop Solution and photometric detection carried out as before.

Following the determination of optimal antibody concentrations, standard curves were constructed and inter- and intra- experimental variation was measured as described below. Samples were also used to determine whether freeze-thawing or keeping plasma samples at 4°C for periods longer than 24 hours would have any detrimental effect on detectable plasma CFH concentrations. This was carried out using the same experimental procedure as described above, on plasma samples obtained from volunteers locally, in triplicate using varying dilutions of plasma i.e. 1:5,000 and 1:7,000.

2.4.3.3 Protocol

Wells of a 96-well microtitre plate (Greiner Bio-One) were coated with 50 µl of 1.2 µg/ml sheep polyclonal anti-CFH antibody (Abcam) diluted in the Coating Buffer. After 2 hours of incubation at room temperature, plates were washed four times using the Wash Buffer. One hundred microlitres of Blocking Solution was used to block unoccupied binding sites. Incubation was carried out at room temperature with shaking, for at least 2 hours. Plates were stored at 4^{o}C overnight after washing the wells four times using the Wash Buffer. EDTA plasma samples were thawed, mixed by inverting, and centrifuged briefly prior to being diluted 1 in 5,000 using the Sample Diluent. Five plasma samples, collected locally and used as internal controls, were pretreated as above. Sample dilutions were carried out using a manual pipette and consisted of a 1 in 100 dilution (10 µl in 1,000 µl) followed by a 1 in 50 dilution (20 µl in 1,000 µl). Eppendorf® tubes (1.5 ml volume) (Eppendorf) were used for diluting samples. Uniform dilution of samples was achieved by leaving the tubes on a rotor (Labinco) to mix for at least 10 minutes at room temperature after each dilution. Doubling dilutions of the 1.5 mg/ml CFH standard were carried out from a starting concentration of 150 ng/ml which was diluted in the Sample Diluent. Fifty microlitres of each sample, standard and internal control samples were each added to the wells of the microtitre plate coated with the capture antibody. After incubation for 2 hours at room temperature, with shaking, plates were washed four times using the Wash Buffer. Fifty microlitres of the 85 ng/ml mouse monoclonal anti-CFH (OX23) detection antibody diluted in the Sample Diluent from the 340 µg/ml stock solution was added to each well and the plate was then incubated for 3 hours at room temperature with shaking. After four washes, 50 µl of the 145 ng/ml HRP-conjugated goat anti-mouse antibody (Biosource UK), diluted in Sample Diluent from the 580 µg/ml stock solution, was added to each well. After 30 minutes of incubation at room temperature with shaking, plates were washed four times using the Wash Buffer. Colour development was achieved after addition of 100 µl TMB Chromogen (Biosource UK) and incubation for 1 hour at room temperature on a shaker. Enzymatic reactions were terminated using 100 µl of the Stop Solution and the absorbance was read at 450 nm using the SkanIt plate reader (Thermo Electron).

Absorbance values obtained using the SkanIt software version 2.1.84 (Thermo Electron) were analysed using Microsoft Excel®. The CFH standard curve was used to calculate sample CFH concentrations using their absorbance values. Internal controls were used to normalise inter-experimental variation as described in Section 2.5.3.

2.4.4 Protein separation by electrophoresis

Protein separation was carried out prior to in-gel staining and western blotting procedures. Two methods of one-dimensional polyacrylamide gel electrophoresis were employed. These consisted of 'home-made' Laemmli gels and pre-cast NuPAGE® Novex Bis-Tris (Invitrogen) gels. Both type of separation technique are described below.

2.4.4.1 Total protein quantitation

Quantitation of the total protein content of each plasma sample was conducted using a spectrophotometer (Jenway) at 280 nm. Quantitation was carried out using 0.1 mg/ml and 1.0 mg/ml of bovine serum albumin (BSA, 98% purity, Sigma) as a standard. The following formula was used to deduce the total protein concentration:

$$\textbf{Concentration (mg/ml)} = A_{280} \textbf{ x Dilution factor x 2.69}$$

A_{280} is the absorbance measured at 280 nm, 2.69 is approximately equal to $1/A_{280}$ of 1 mg/ml of BSA protein standard. Serial 10-fold dilutions of the sample were prepared using Millipore® dH_2O, starting with a 1 in 10 dilution.

2.4.4.2 Protein separation using Laemmli gels

Denaturing polyacrylamide gel electrophoresis (PAGE) was carried using Hoefer® SE 600 Vertical Slab units. After cleaning the glass parts with 70% ethanol, 10% sodium dodecyl sulphate (SDS) resolving gels were prepared (**Table 2.7**) and poured to approximately 1 cm below the position of the comb. Gels were then overlaid with approximately 1 ml of isobutanol saturated with dH_2O, at 1:1 ratio, to prevent the formation of air bubbles. After gel setting, the isobutanol was discarded and the stacking gel (**Table 2.7**) was poured on top of the resolving gel and left to set.

	Resolving gel (10%)	Stacking gel (4%)
H_2O (Millipore)	7.04 ml	13.35 ml
Tris-HCl (Sigma) buffer	5.00 ml	2.40 ml
40% Acrylamide (BioRad)	4.86 ml	1.80 ml
2% Bis-acrylamide (BioRad)	2.68 ml	975 μl
20% SDS (Fisher Scientific)	0.20 ml	188 μl
10% Ammonium persulphate (Sigma)	0.20 ml	189 μl
Tetramethylethylenediamine (Invitrogen)	0.02 ml	37.5 μl

Tris-HCl buffer was used at 1.5 M pH 8.8 and 1.0 M pH 6.8 in the resolving and stacking gels respectively. Volumes and concentrations are as recommended by Sambrook and Russell, 1989.

Table 2.7 Chemical composition of the Laemmli gels used in one-dimensional protein separation.

Protein samples were incubated with 1x SDS Loading Buffer for 10 minutes at 70°C. Gels were run in 1x SDS Running Buffer after loading the denatured samples and a broad range protein marker (BioRad). Electrophoresis was conducted at 50 mV overnight or until the dye front reached the bottom of the gel.

2.4.4.3 Protein separation using pre-cast gels

One-dimensional pre-cast NuPAGE® Novex Bis-Tris (Invitrogen) gels were also used for protein separation. Prior to denaturation and protein reduction, plasma samples were diluted 1 in 10 using distilled water (dH_2O, Millipore), in order to reduce the complexity of plasma. Samples were incubated with 10 μl of 4x NuPAGE® sample buffer (Invitrogen), 5 μl of 50 mM dithiothreitol (DTT) reducing

agent (Invitrogen) and dH_2O (Millipore) to bring the volume up to 40 µl. Samples were denatured at 70°C for 10 minutes and loaded on 4-12% NuPAGE® Novex Bis-Tris gels (Invitrogen) immersed in 1x SDS Running Buffer. The running buffer was prepared by adding 40 ml of 20x NuPAGE® 3-(N-morpholino) propane sulfonic acid (MOPS) SDS Running Buffer (Invitrogen) to 760 ml of dH_2O (Millipore). A "SeeBlue" Plus 2 prestained molecular weight marker (Invitrogen) was also loaded and electrophoresis was conducted at 200 volts for approximately 45 minutes.

2.4.5 Semi-dry transfer of proteins to membranes

Transfer of proteins separated using either of the transfer methods mentioned above was carried out using a semi-dry transfer device (BioRad). Proteins were transferred to a Hybond™-C Extra nitrocellulose membrane (Amersham Biosciences) by firstly soaking in methanol (Fisher Scientific), then transferring to dH_2O (Millipore) and finally transferring into Transfer Buffer. Six pieces of thick filter paper, the same size as the membrane, were also immersed in Transfer Buffer. On the lower negative electrode of the transfer device, three pieces of filter paper, the membrane, the gel containing the separated proteins and finally another three pieces of filter paper were layered in order. Air bubbles were removed using a plastic roller and transfer was conducted at or greater than 1.2 mA/cm^2 at a maximum of 25 volts for 60 minutes.

2.4.6 Western blotting

The following is a general protocol used for the immunodetection of proteins after semi-dry transfer. Details of antibodies and their concentrations are shown separately for individual experiments, in the Results chapters. Membranes were blocked for 1 hour using an excess of phosphate buffered saline (PBS)-Tween® including 5% dried bovine skimmed milk (Premier International Foods), with gentle shaking at room temperature. Incubation with the primary antibody, diluted in PBS-Tween® containing 5% dried milk, was carried out in a sealed plastic bag with gentle shaking over night at 4°C. After three 10 minute washes in PBS-Tween® membranes were incubated with the appropriate HRP-conjugated secondary

antibody, diluted in PBS-Tween® containing 5% dried milk. After incubation on the rotor for 1 hour at room temperature, using a sealed bag, membranes were washed three times as before.

Detection of the secondary antibody was carried out by incubating the membranes with enhanced chemiluminescence (ECL) western blotting detection reagents (Amersham Biosciences), according to the manufacturer's instructions. Excess detection reagents were drained off and the membrane placed in an X-ray film exposure cassette with a sheet of Kodak™ X-Omat AR imaging film placed on top of it. Following sufficient time for exposure, a phosphorimager X-ray machine (Konica Minolta) was used for the development of the film.

2.4.7 SDS-PAGE gel staining

Two methods of blue and silver staining were used for the visualisation of proteins separated by SDS-PAGE. The blue staining was carried out prior to the IEC mass spectrometry experiments, as described later. The silver staining was employed as a more sensitive method of visualising polypeptides after electrophoresis.

2.4.7.1 SimplyBlue™ staining

SimplyBlue™ SafeStain (Invitrogen) was used for the staining of plasma proteins separated on NuPAGE® Novex Bis-Tris gels (Invitrogen). Staining was carried out in order to deduce the approximate position of the full length CFH protein band after gel electrophoresis. SimplyBlue™ also provides the right level of sensitivity for differentiating between separated plasma proteins.

After electrophoretic separation of plasma proteins, the SDS-PAGE gels were stained with 20 ml of SimplyBlue™ SafeStain (Invitrogen). Incubation was carried out overnight on a shaking rotor, followed by destaining in MilliQ™ dH2O (Millipore) for 60 minutes, to reduce the background stain.

2.4.7.2 Silver staining

Staining of polypeptides with a silver stain after gel electrophoresis is a process that relies on the differential reduction of silver ions that are bound to the side chains of amino acids (Switzer *et al*., 1979; Oakley *et al*., 1980; Merril *et al*., 1984). Staining with silver nitrate is advantageous over ammoniacal silver solutions in that silver nitrate solutions are easier to prepare and do not generate potentially explosive by-products. Furthermore, the detection limit of this type of staining is as little as 0.1-1.0 ng of protein.

Silver staining was carried out to assess the purity of CFH standard used in the assay of complement factor H. It was also used to confirm the depletion of high abundant proteins from plasma after using the multiple affinity removal system.

SDS-PAGE gels containing the separated proteins were incubated in 5 gel volumes of the Fixing Solution for 4 hours with gentle shaking. Gels were then incubated twice in 5 gel volumes of 30% ethanol (Fisher Scientific) for 30 minutes and followed by two incubations for 10 minutes in 10 gel volumes of dH_2O (Millipore). Staining was carried out for 30 minutes using 5 gel volumes of silver nitrate Staining Solution, diluted 200 times from a 20% stock, with gentle shaking. The Staining Solution was removed by washing both sides of the gel for 20 seconds under a stream of dH_2O (Millipore). Five gel volumes of the Developing Solution were then added to each gel and development was carried out with gentle agitation at room temperature. Finally the reaction was quenched by washing the gel in 1% acetic acid (Fisher Scientific) for 5 minutes, followed by three washes in dH_2O (Millipore) for 10 minutes. The stained gel was then dried and photographed.

2.4.8 Proteomic and mass spectrometry methods

Mass spectrometry techniques using immuno-depletion and magnetic bead separation and MALDI-TOF were developed and optimised in collaboration with Dr. A. Henzing and Dr. J. Creanor (University of Edinburgh). Capillary electrophoresis and LC-MS techniques were optimised with help from Dr. C. L. MacKay (University of Edinburgh). Ion echange chromatography experiments and the follow-up nLC-

MS/MS procedures were conducted as a service by Mr. D. Lamont (University of Dundee).

2.4.8.1 Immunodepletion of high abundance plasma proteins

2.4.8.1.1 Background

The multiple affinity removal system allows for the immunodepletion of unwanted, high abundance proteins from human plasma or CSF samples (Pieper *et al*., 2003). It consists of a multiple affinity column containing polyclonal antibodies to six high abundance plasma proteins immobilised on sepharose beads (albumin, immunoglobulin G, alpha-1 antitrypsin, immunoglobulin A, transferrin and haptoglobin). The multiple affinity columns are widely used either alone or in combination with other treatment methods, including magnetic bead separation prior to protein analysis (e.g. MALDI-TOF, LC-MS and iTRAQ™).

2.4.8.1.2 Protocol

Ten microlitres of crude human plasma was diluted 20 times in 190 µl of "Buffer A" (Agilent Technologies). The diluted solution was filtered through a 0.22 µm spin filter (Agilent Technologies) in a centrifuge at 2,795 *g* for 5 minutes. In the meantime, 4 ml of "Buffer A" (Agilent Technologies) was slowly added to the filter resin using a syringe to remove any trapped air bubbles and equilibrate the Multiple Affinity Removal Spin Cartridge (Agilent Technologies). Excess "Buffer A" was then removed from the top of the cartridge. The first fraction, containing most of the low abundance proteins (F1), was obtained by centrifugation of 200 µl filtered plasma for 90 seconds at 100 g. The remainder of the low abundance proteins bound to the filter were collected in the same tube by filtering 400 µl of Buffer A for 2.5 minutes at 100 g. The second fraction of the low abundance proteins (F2) was obtained by adding 400 µl of "Buffer A" to the top of the resin bed followed by centrifugation for 2.5 minutes at 100 g. The fraction containing the more abundant plasma proteins bound to the cartridge was eluted by placing the spin column in a new 1.5 ml tube (Eppendorf) followed by addition of 2 ml of "Buffer B" (Agilent Technologies) using a syringe. Finally, 4 ml of "Buffer A" was used to re-

equilibrate the cartridge according to the manufacturer's instructions. F1, F2 and the eluted fraction were then stored at 4^{o}C until further use.

2.4.8.2 Hydrophobic interaction chromatography

2.4.8.2.1 Background

One of the most recent advances in the separation of specific peptides from a mixture of proteins and peptides has been the development of magnetic beads. Peptides and proteins are captured from complex biological material such as plasma, urine or cerebrospinal fluid and concentrated on magnetic micro-particles or beads. Once the bound molecules have been eluted from the magnetic beads, they can be analysed by MALDI-TOF MS without further purification. This technique allows for the simultaneous detection of large numbers of plasma proteins ranging in size from 0.8 to 15 kDa (Villanueva *et al*., 2004; Zhang *et al*., 2004b). Six different surface properties including hydrophobic interaction chromatography (C3, C8 and C18), immobilised metal ion affinity chromatography, weak anion exchange chromatography and weak cation exchange chromatography, can often be used. Varying the functionality of the beads allows the enrichment of different peptides and proteins, thus increasing the number of biomarkers that can be found (Bruker Daltonics®).

2.4.8.2.2 Protocol

Three magnetic beads for hydrophobic interaction chromatography were used, which included magnetic bead high interaction chromatography (MB-HIC) C3, C8 and C18 resins (Bruker Daltonics). Beads were shaken gently at least 20 times prior to each use. F1 and F2 fractions were merged after immunodepletion of high abundance plasma proteins. Ten microlitres of MB-HIC Binding Solution (Bruker Daltonics) was added to 100 µl of the F1/F2 fractions in a thin 0.2 ml PCR tube (Agilent Technologies). Five microlitres of the magnetic bead mixture was added to the mixture and mixed gently by pipetting five times. The mixture was allowed to equilibrate at room temperature for 1 minute, before being placed into a magnetic bead separator (MBS) (Bruker Daltonics) for 20 seconds to achieve separation of the

beads from the sample. The supernatant was collected with a pipette and stored for further analysis, while avoiding any contact between the beads and the pipette. After transferring the tube from the MBS to an alternative device, 100 µl of MB-HIC Wash Solution (Bruker Daltonics) was added to the sample and the tubes were placed back into the MBS. After 20 seconds the supernatant was disposed of gently and the beads were washed 3 times using the Wash Solution. Elution of the protein and peptides bound to the beads was carried out by incubation with 10 µl of 50% acetonitrile solution (Sigma) for 1 minute at room temperature. After separation using the MBS for 30 seconds, the supernatant containing the eluted peptides and proteins was removed and stored at 4^{o}C for further use.

2.4.8.3 Pretreatment with ZipTip™

Concentrating, desalting and purification of crude plasma samples was carried out using the ZipTip™ technology. ZipTip™ is a 10 µl pipette tip with a bed of chromatography media fixed at its end. ZipTip™ is capable of purifying and concentrating femtomoles to picomoles of protein and peptides. Two types of C18 and C4 ZipTip™ tips were used, targeted at the low and the high molecular weight ranges respectively.

Tips were first equilibrated using the wetting solution (50% acetonitrile in TFA in Milli-Q® dH_2O) followed by washing in 0.1% TFA in Milli-Q® dH_2O. Interaction between the diluted samples and the chromatography bed was carried out by pipetting up and down gently a few times. Elution of the bound fraction of plasma was performed using 50% acetonitrile in Milli-Q® dH_2O. Samples were eluted directly onto a MALDI plate for MS analysis.

2.4.8.4 Ultrafiltration of plasma

2.4.8.4.1 Background

Ultrafiltration (UF) through membranes with specific molecular weight cut-offs can be used to prepare low-molecular-weight fractions for MS and biomarker analysis. UF is in many ways better than alternative separation methods as it can be used to prepare larger proteins (>10 kDa) for electrophoresis and to simultaneously separate

smaller polypeptides (<10 kDa) for MS analysis (Chernokalskaya *et al.*, 2004). Other advantages include speed and protein recovery. Peptides present in the ultrafiltrate are often analysed after being concentrated and desalted by solid phase extraction on hydrophobic interaction chromatography resins (e.g. C18 beads).

2.4.8.4.2 Protocol

This protocol was kindly provided by Elena Chernokalskaya (Millipore) for high throughput ultrafiltration of plasma peptides for MALDI TOF MS. Three to five hundred microlitres of crude plasma was filtered using 10 kD molecular weight cut-off filters (VivaScience). Three to five hundred microlitres of control plasma samples diluted 1:1 in 10 mM Tris-HCl pH 7.5 was also filtered. Centrifugation was carried out for 30 minutes at 1,006 *g*. Ten microlitres of the filtrate was acidified with 5 µl of 1% trifluroacetic acid (TFA) pH 0.9 (Sigma). Samples were desalted and concentrated using µC18 ZipTip™ pipette tips (Millipore) as before.

2.4.8.5 MALDI-TOF MS analysis

One microlitre of either the product of ultrafiltration or MB-HIC was co-eluted with 1-2 µl of 5 mg/ml alpha-cyano-4-hydroxy cinnamic acid matrix, onto a MALDI target plate (Applied Biosystems). Samples were left to dry in a fume hood. Spectra of mass to charge ratio were obtained and recorded using a MALDI-TOF MS Voyager-DE™ STR (Applied Biosystems) in a linear mode. The laser intensity was altered between 2,100-2,500 µJ/pulse depending on the window size of the proteins/peptides of interest i.e. lower laser intensity for smaller protein size. The results were analysed using the software Data Explorer version 4.0 (Applied Biosystems).

2.4.8.6 Capillary electrophoresis coupled with MS

The following protocol is a modified version of the capillary electrophoresis (CE) technique, which was used in conjunction with electrospray ionisation (ESI) coupled with tandem Fourier transform ion cyclotron resonance (FT-ICR) MS (Chalmers *et al*., 2005). An Agilent 3D capillary electrophoresis system (Agilent Technology)

was used in the analysis. Pretreatment of the 90 cm x 75 μm capillary (Agilent Technologies) was carried out once a day prior to the start of the experiment and consisted of rinsing the capillary with MilliQ™ dH_2O (Millipore) for 5 minutes, followed by 1 M NaOH (Sigma) for 15 minutes, dH_2O (Millipore) for 15 minutes, ammonia (BDH) for 25 minutes and dH_2O for 15 minutes, using a pressure of 25 psi.

The experimental procedure consisted of electrophoretic separation of proteins at room temperature using the Running Buffer at 30 kV for 80 minutes, using a flow rate of 500 nl/min. After 10 minutes, 1 M ammonia (BDH) was injected for 7 seconds followed by injection of the protein sample diluted in the Running Buffer, at a positive pressure of 1 psi. Two molar formic acid (Sigma) was finally injected into the capillary for 5 seconds. Pressure was applied at 0.1 psi after 40 minutes and raised by 0.1 psi every 2 minutes until reaching 0.5 psi. Data acquisition was carried out at 280 nm when in the UV absorbance mode and the results were analysed using the Agilent ChemStation software (Agilent Technology).

2.4.8.6.1 CE analysis of single protein sample

Initially, three proteins consisting of ubiquitin (8.5 kg/mol, 90% purified, salt-free lyophilised powder from bovine erythrocytes, Sigma), myoglobin (16.7 kg/mol, 95-100% salt-free, lyophilised powder from equine skeletal muscle, Sigma) and BSA (66.4 kg/mol, 99% lyophilised powder, Sigma) were used separately at two different concentrations of 10 μg/ml and 50 μg/ml. Protein samples were diluted in the CE Running Buffer. Negative controls, consisting of the Running Buffer alone, were also included in each run. The capillary was washed after each run, using the protocol described before.

Furthermore, some replicates of the negative control samples were found to contain the same peaks at approximately the same time points as those containing ubiquitin, myoglobin and BSA, and high levels of background noise. The troubleshooting procedures, described in Section 2.4.8.6.2 were employed to eliminate the problem of high background noise and contaminating peaks detected in the CE spectra. Following the troubleshooting procedure which consisted of filtering all buffers, avoiding air bubbles and dissolving the protein samples in the Running Buffer thoroughly, the resulting CE spectra consisted mainly of low levels of

background noise and no peaks of ubiquitin, myoglobin or BSA were detected in the control samples.

If CE was coupled with ESI-FT-ICR, MS detection was performed on a 12-T quadrupole system APEX-QE (Bruker Daltonics, Germany) equipped with a fourier transform mass spectrometer using an electrospray ionization interface (Agilent Technologies) and a CE probe. The sheath liquid consisted of the CE Running Buffer and was delivered by a syringe pump with a flow rate of 150 µl/h. The drying gas flow rate was 1.4 dm^3/min (nitrogen 5.0 at a temperature of 150^oC), a nebulising gas was not applied.

2.4.8.6.2 CE troubleshooting

The possibility of contaminated buffers, interfering air bubbles and precipitated samples was investigated as follows. All buffers were filtered though 0.2 µm filters (Millipore) in order to exclude precipitants and contaminants in buffers. Precipitation of protein samples was prevented by verifying that sample components were sufficiently soluble in the buffer. Care was also taken to not introduce air bubbles into buffers while pipetting.

Low signal to noise ratio in CE can be the result of occluded optical slits in capillary interface or ageing Deuterium UV lamp. In order to overcome the problem of high background noise in CE spectra, the optical slit was regularly cleaned with methanol or water. The Deuterium lamp was also replaced with a new lamp (Agilent Technologies) and calibrated.

2.5 Statistical analyses

2.5.1 Predictive power calculation

The power of a study is the probability of successfully detecting an effect of a particular size and principally depends on effect size and sample size (Sham *et al*., 2000). If β is the probability of a false-negative or type II error, then power is calculated as $1 - \beta$. The Genetic Power Calculator (http://pngu.mgh.harvard.edu/~purcell/gpc) used for the estimation of power was based on the above formula (Purcell *et al*., 2003).

Factors that contribute to effect size are typically unknown, therefore assumptions were made about the magnitude of effect, sample size (*N*) and required level of statistical significance, α, defined as the false-positive, or type I error rate. Levels of linkage disequilibrium (LD) between marker SNP and QTL of interest and allele frequencies were also assumed in order to estimate study power. The number of unrelated individuals present in the Vis data set (*N*) was set at 500, α was set at 0.01 and pairwise LD (r^2) was assumed to be 0.80 between the marker and QTL of interest. Power was then estimated for varying levels of effect size and allele frequencies, as shown in the Results chapter. A variance components model was used for the estimation of power in a test of association based on differences in means, given the genotypes at the candidate locus (Sham *et al*., 2000). The formulae used in the power estimation make use of the relationship between the average values of the test statistics and genetic parameters such as the proportion of variance accounted for by the QTL (between 0 and 1) and the degree of linkage disequilibrium between QTL and marker locus.

2.5.2 Empirical analysis of power using plasma lipids

The *APOE* allelic model was used to test for evidence of additivity or dominance associated with each of the *APOE*E2*, *APOE*E3* and *APOE*E4* alleles. For an additive model, *APOE* genotypes were coded as 0, 1 or 2 based on the number of copies of the allele to be examined. For the dominant model, *APOE* genotypes were coded as either 0 or 1 according to the presence or absence of the allele to be examined. Changes in the mean plasma levels associated with *APOE* genotypes were assessed using a binary coding system, after converting each genotype class into a separate variable. Genotypes were coded as 1 or 0 depending on the presence or absence of the particular genotype in that variable class. The mean QT was then applied to the most common class (*APOE*E3* homozygous genotype). The effects of sex, age, BMI and smoking on plasma lipid concentrations were also tested in each set of analyses. The effect of each allele or genotype was measured in terms of the proportion of the QT variance explained by the covariate (R^2) and the direction of the effect was shown by β, the regression coefficient of the linear correlation

between the trait and the genetic variant. Negative values of β indicate a reduction in the mean trait value.

2.5.3 Normalising inter-experimental variation in ELISA of CFH and Aβ

Variation between repeated assays is a result of experimental technique, environmental conditions and performance characteristics of the assay method. Therefore, a set of appropriate control samples, assayed repeatedly over time, was used to correct for inter-assay variation observed in ELISA.

Normalisation of each of the replicated measurements was carried out as follows:

$$\textbf{Normalised measurement} = (x / y) \times \textbf{raw measurement}$$

in which x is the mean of the means of the five internal triplicate control assays measured over the course of all assays (> 20) and y is the mean of the means of five internal controls for one particular assay.

2.5.4 Detection of outliers in assays of CFH and Aβ

Quantitative determination of plasma CFH, Aβ40 and Aβ42 concentrations was carried in triplicate. Two methods of detecting large variation between measurements were employed. First was the use of the coefficient of variation (CV), calculated as follows:

$$CV\ (\%) = (\sigma / \mu) \times 100$$

σ is the standard deviation of the mean (μ) of the three replicates from the same sample. The standard threshold of significance for CV is 5%. The CV percentage provides a reliable estimate of the variation observed between measurements of the same sample.

A second statistical technique for dealing with outliers is referred to as Grubb's test (Taylor 1990; Iglewicz & Hoaglin 1993; Barnett & Lewis 1994; Aggarwal *et al*., 2005), also called the extreme studentised deviate (ESD) method. Elimination of outliers resulted in more robust mean measurements with lower CVs.

The deviation of an outlier from a set of values is quantified using the formula:

$$Z = (\mu - x) / \sigma$$

in which Z is the difference between the outlier and the mean (μ) of all replicated measurements, including the outlier (x), divided by the standard deviation (σ). A 5% threshold for significance was set, which meant that 5% of the values in a set of samples, assuming a Gaussian distribution, were more than 1.96 (Z) standard deviations from the mean. The value of Z obtained above was compared to a table of critical values of Z for a given number of samples. The critical value of Z for $N = 3$ was 1.15. The two-tailed P-value was estimated for each measurement within a set of $N = 3$, by firstly calculating the student t distribution using the following formula:

$$t = \sqrt{[N (N - 2) Z^2] / [(N - 1)^2 - NZ^2]}$$

Using the student t distribution and N-2 degrees of freedom, the approximate two tailed P value was estimated after taking the number of samples into consideration. The value of P is most accurate for larger sample sizes and is over estimated for small N.

2.5.5 Testing for normality

The Kolmogorov-Smirnov goodness-of-fit test was used to check the normality of the sample distributions using the statistical software SPSS v.12. The test compares the set of values in the sample to a normally distributed set of values with the same mean and standard deviation. A non-significant test result ($P > 0.01$) indicates that

the distribution of the sample is not significantly different from a normal distribution i.e. the sample distribution is normal.

2.5.6 Quantile normalisation

Variance components and regression analyses are sensitive to kurtosis and skewness in the trait distribution. Failure to normalise such distributions would result in an increased probability of obtaining false positive results.

Quantile normalisation (Pilia *et al*., 2006; Sokal & Rohlf, 2000) was carried out by first ranking the observations and then matching the percentile of each observation to the corresponding percentile in a standard normal distribution. The resulting percentiles were then used to obtain the corresponding normal equivalent deviates (NED) or Z-scores for each observation (**Figure 2.5**). When ties were present, percentiles were averaged across all ties.

2.5.7 Linear regression analyses

General linear models were used in a number of analyses to estimate the coefficients of the linear regression, involving one or more independent variables that best predict the value of the dependent variable. The following applications were all variations of the general linear model, and are explained in detail.

2.5.7.1 Multiple regression

The general linear model implemented in the software SPSS version 12.0 was used to assess the significance of effects of well-characterised continuous covariates and /or categorical cofactors on the quantitative trait outcome. The effect size is estimated using the linear equation:

$$y_i = \mu + \sum_j \beta_j c_{ji}$$

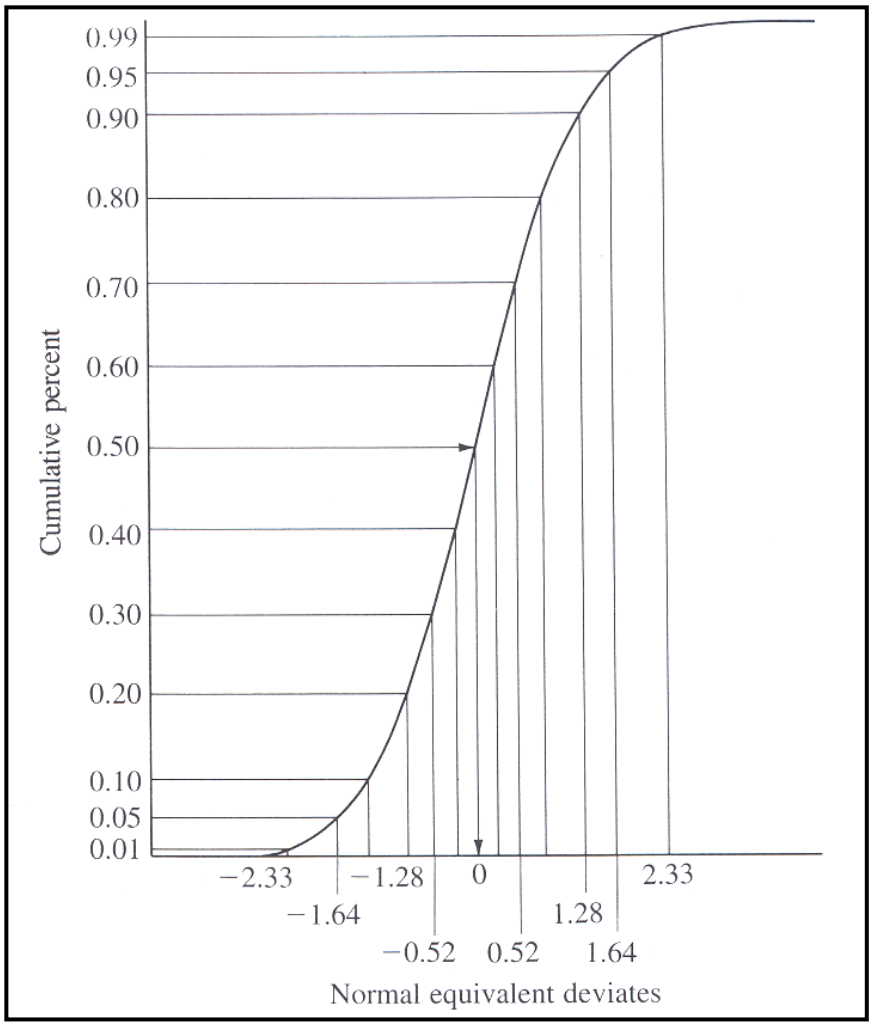


Figure 2.5 Transformation of cumulative percentages into normal equivalent deviates.

Cumulative percentiles from a standard normal distribution were transformed into a standard deviation scale, called normal equivalent deviates (NED). Ranked percentiles of observed data measurements were replaced by a z-score (NED). Note that there are no 0% or 100% points since the normal frequency distribution extends from negative to positive infinity. Diagram copied from Sokal & Rohlf, 2000.

where y_i is the phenotype of the ith individual, μ is the mean value of the trait in the sample, c_{ij} is the value of the jth predictor for the individual i (e.g. age) and β_j is an estimate of the magnitude of effect associated with the predictor. β is the regression coefficient of the general linear model and represents the change in the quantitative trait associated with a unit change in the predictor, and therefore takes the same unit as the outcome variable. This also holds true for assessing the effect associated with a biallelic marker under an additive model as marker genotypes are coded 0, 1 or 2 based on the number of copies of the minor allele.

A t-statistic was used to determine whether values of β differed significantly from zero, the null hypothesis value of β,

$$t = (\beta_{observed} - \beta_{expected}) / SE_{\beta}$$

with $\beta_{expected}$, the expected value of β under the null hypothesis and SE_{β}, the standard error associated with β.

The linear regression methodology was used to estimate the proportion of the phenotypic variance explained by the components of the model (R^2) and to assess the goodness of fit of the model (F statistic). The multi-locus model used in the univariate analysis of plasma CFH was derived using this methodology. R^2 is calculated based on the amount of variance in the outcome explained by the model, relative to the amount of variation present in the first place:

$$R^2 = SS_M / SS_T$$

SS_T represents the sum of squares of the differences between each of the observed data and the mean value of the outcome. SS_M is the difference between the SS_T and the residual sum of squares (SS_R) and symbolises the sum of differences between each of the observed data and the regression line.

The goodness of fit of the model is assessed by the least squares method which favours a model with minimal residual sum of squares. The F-test is based on

the ratio of improvement due to the model, and the difference between the model and the observed data:

$$\mathbf{F = MS_M / MS_R}$$

Here, MS_M represents the mean sums of square of the model and MS_R shows the residual mean sums of squares. In order to calculate the mean sums of squares, the sums of squares were divided by the degrees of freedom. In the case of MS_M this was the number of variables in the model and for MS_R it was the number of observations minus the number of parameters being estimated.

2.5.7.2 Stepwise linear regression analysis of markers

A 'backward' linear regression analysis was carried out to obtain the multimarker model best associated with levels of plasma CFH, using the software SPSS version 12.0. The analysis consisted of placing all predictors in the model and assessing the contribution of each one by calculating the significance value of the t-test for each predictor. Each predictor was tested against a removal criterion (explained in Results Chapter 7) which would result in the predictor being kept or dropped. The model was then re-estimated for the remaining predictors, and the contribution of each was assessed as before.

2.5.7.2.1 Detecting collinearity among predictors

An important assumption in any linear regression analysis is the absence of collinearity between subjects. Collinearity exists when there is a strong correlation between two or more predictors in the regression model.

The statistics of collinearity estimates are very complex. In short, collinearity affects the validity of regression analysis in three ways. Firstly, it limits the size of the correlation coefficient R as uncorrelated predictors account for different (unique) variances in the outcome. Hence when both predictors are included in the model, R will be substantially larger than when correlated predictors which account for the same variance are included in the model. Secondly, collinearity between predictors

makes it difficult to assess the individual importance of a predictor as highly correlated predictors account for similar variances in the outcome. Thirdly, as collinearity between two or more predictors of a variable increases, so do the standard errors of the β coefficients, which in turn affects whether these coefficients are found to be statistically significant.

Two estimates were used to represent collinearity between predictors, variance inflation factor (VIF) and tolerance (1/VIF),

$$tolerance = 1 - R^2, \text{ VIF} = 1/tolerance$$

in which R^2 represents the squared correlation between variables. VIF indicates whether a predictor has a strong linear relationship with the other predictor(s). VIF values greater than 10 were taken as indicative of redundant predictors (Myers, 1990). Similarly, values of the tolerance statistic below 0.1 were indicative of collinearity (Menard, 1995).

2.5.7.3 Multi-locus association analysis

Phasing of a given group of SNPs and haplotype association analysis was carried out using the conditional haplotype association test, implemented in the software package PLINK (Purcell *et al.*, 2007). A default 'omnibus' test was used to compare the null and alternate models. The alternate model was based on the assumption that each haplotype had a unique effect and the null assumed that no haplotypes had any different effects. Coefficients of regression were estimated for each phased haplotype with respect to a reference haplotype, chosen arbitrarily. An F-test was used to compare the models for H haplotypes with H-1 degrees of freedom.

2.5.7.4 The linear mixed model

To take into account the non-independence of the outcome measures due to relatedness of the participants, a linear mixed model was used with a polygenic "background genetic" effect fitted as a random effect. Random effects are those that vary in the population, have means of zero, are characterised by their variance and

affect the covariance of the data. The mixed linear model is a modification of the multiple regression model,

$$y_i = \mu + \sum_j \beta_j c_{ji} + G_i + e_i$$

This model adjusted the quantitative trait for a random additive polygenic effect, G_i and a residual effect, e_i, with an additive genetic variance σ^2_A, and a residual variance σ^2_e, respectively. Variance parameters and fixed effects (those with well-characterised effects, such as sex that do not vary across individuals) were estimated using the maximum likelihood method in SOLAR (freely available, Almasy & Blangero, 1998) or by the restricted maximum likelihood method in ASReml (commercial software under license, VSN International 2002).

2.5.7.5 Estimates of heritability

Narrow-sense heritabilities were used to estimate the proportion of the variance of the quantitative trait due to additive genetic factors, environmental factors and covariates using the software package SOLAR (Almasy & Blangero, 1998). The linear mixed model used information on the covariance among relatives to estimate narrow-sense heritabilities (h^2),

$$h^2 = \sigma^2_A / \sigma^2_P$$

which estimates the proportion of the total phenotypic variance (σ^2_P) due to the additive genetic variance (σ^2_A).

2.5.7.6 Genome-wide association analysis

Genome-wide tests of association were carried out using a standard linear regression model of phenotype on allele dosage (Boerwinkle *et al*., 1986) implemented in the genome-wide association tool set PLINK (Purcell *et al*., 2007). Population

stratification as a result of the clustering of the Croatian dataset into two sub-groups (roughly corresponding to the two villages) was accounted for, using a permutation approach (Purcell *et al*., 2007). Covariate effects were also accounted for prior to the association analysis (described below). Data from 966 individuals was used in the test of genome-wide association, 58% of whom were members of 122 families. Seventeen individuals had been removed due to low genotyping rate (<90%) and three due to non-European ethnicity.

The genome-wide rapid association using mixed model and regression (GRAMMAR) method was used to take into account the relatedness of the participants and covariate effects (Aulchenko *et al*., 2007). First, the data was analysed by Dr. V. Vitart using the linear mixed model. The residuals (y^*) from this analysis were obtained by:

$$y_i^* = y - (\mu + \sum_j \beta_j c_{ji} + G_i) = \hat{e}_i$$

The residuals adjusted for relatedness, covariate and fixed effects were then used as the dependent variable in a linear regression analysis for estimating the effect of each SNP,

$$\hat{e}_i = \mu + kg_i + e_i$$

where e_i is the vector of residuals (obtained above), μ is the mean quantitative trait, g is the vector of genotypes (additive or dominance model) at the marker under study, k is the marker genotype effect and e_i is the vector of random residuals.

2.5.7.6.1 Correction for multiple testing

Correction for multiple testing was carried out using the Bonferroni correction with the significance threshold of $P < 1.6E\text{-}07$, equivalent to a 5% genome-wide significance (α′). The following formula was used to obtain a significance threshold after correction for 308,140 independent tests (A total of 317,503 SNPs were

genotyped, of which 8,984 were removed due to low genotyping call rates (< 90%) and 379 were removed due to low minor allele frequency (<1%) leaving 308,140 SNPs):

$$\alpha' = 1 - (1 - \alpha)^{1/308140}$$

in which α would be equivalent to a 0.05 level of significance and α' is the threshold of significance after correction for multiple testing ($\alpha' \approx$ 1.7E-07).

2.5.8 Genome-wide linkage analysis

The pedigree-based variance component linkage methodology that maximises the use of all available phenotypic information was used in a genome-wide linkage analysis implemented in the software package SOLAR (Almasy & Blangero, 1998). This method estimates the amount of variance in a quantitative trait due to a particular genetic locus (σ_q^2), residual genetic factors (σ_g^2) and individual-specific, random environmental factors (σ_e^2), based on the phenotypic covariance between arbitrary relative pairs. An additional random QTL effect is fitted in the linear mixed model and tests for significance approximately every 5 cM along the genome. The variance-covariance matrix takes the form:

$$\Omega = \sigma^2_q \Pi + 2\sigma^2_g \Phi + \sigma^2_e I$$

where Π is the matrix whose elements provide the predicted proportion of alleles (0, 1 or 2) that each pair of individuals share IBD at a QTL linked to a genetic marker locus. Φ is the kinship matrix describing polygenic factors and I is the identity matrix describing sporadic environmental factors. IBD probabilities were calculated by Dr. V. Vitart from the STR genotype information using the statistical software Loki (Heath, 1997).

The model tested for multipoint linkage by comparing the likelihood of the restricted model in which the variance due to the QTL (σ^2_q) is fixed to 0, to that of a model where σ^2_q is estimated. The difference between the two $\log_{10}$ likelihoods produced a LOD score equivalent to the classical LOD score of linkage analysis.

2.5.9 Estimation of Hardy-Weinberg equilibrium

Hardy-Weinberg equilibrium (HWE) describes a mathematical relationship between alleles and genotypic frequencies. If p is the frequency of the allele A, then the terms p^2, 2pq and q^2 will estimate the frequencies of AA, Aa and aa respectively. A number of assumptions are made for the test. Firstly, that the population is large enough in size so that allele frequencies are not subject to change simply because of the sampling which leads to random genetic drift. Secondly that the allele frequencies have not changed from one generation to the next and finally that mutation, migration and natural selection are negligible.

A Fisher's exact test of HWE (Wigginton *et al*., 2005), implemented in the genome-wide association tool set PLINK (Purcell *et al*., 2007) was used. The test is conditional on the allele frequencies, *p* and q, and the hypothesis of Hardy-Weinberg proportions is rejected if the frequency of heterozygous genotypes is too large or too small. The conditional probabilities for the heterozygotes given the allele frequencies is calculated as follows:

$$prob[n_{12}|n_1] = \frac{\binom{n}{n_{11},n_{12},n_{22}}}{\binom{2n}{n_1}} 2^{n_{12}},$$

where, n_{11}, n_{12} and n_{22} are the observed frequencies of the three genotypes,and the frequency of the n_1 allele is calculated as $n_1 = 2n_{11} + n_{12}$. The sample size is denoted by *n*. *P*-values representing the probability of deviations from the Hardy-Weinberg equilibrium were calculated.

2.5.10 Analysis of linkage disequilibrium

The squared correlation, r^2 was used as the measure of linkage disequilibrium (Crow & Kimura, 1970). It has been shown that r^2 can be related to population genetics parameters as it is inversely proportional to the effective population size ($r^2 \sim (1 + 4N\theta)^{-1}$, in which θ is the recombination fraction between the two loci) and directly reflects the statistical correlation between the alleles at two loci (Wall & Pritchard,

2003). Furthermore, r^2 is not adjusted to loci having different allele frequencies and exhibits more reliable sampling properties since it is not biased upward in small sample sizes and for low allele frequencies (Shifman *et al*., 2003). The r^2 measure is a normalised estimate of D, the first measure of linkage disequilibrium introduced, which is highly sensitive to differences in allele frequency (Robbins, 1918).

The software package Haploview (Barrette *et al*., 2005) was used to estimate pairwise r^2 for biallelic markers,

$$\mathbf{r^2 = D^2 / (p_1q_1\ p_2q_2)}$$

where p and q represent the two alleles at a biallelic locus. *D* estimates the difference between the observed frequency of co-occurrence of both alleles on the same chromosome and the expected frequency of co-occurrence under linkage equilibrium,

$$\boldsymbol{D = h_{11} - p_1q_1}$$

where h_{11} is the frequency of the haplotype containing the two alleles, namely A_1 and B_1, and p_1q_1 is the product of frequencies of A_1 and B_1. The magnitude of *D* is proportional to $(1 - \theta)^g$, where θ and g are, respectively, the recombination fraction between the loci and the number of generations since the mutational event responsible for the most recent polymorphism (Sham *et al*., 2000).

2.5.11 Bioinformatic resources

The following online databases were used in the bioinformatic analysis of results.

- **International HapMap Project** (www.hapmap.org)
  Data Rel 21A/PhaseII, Jan 2007 based on the NCBI Build 35 assembly

- **The National Centre for Biotechnology Information**
  (www.ncbi.nlm.nih.gov)

- **UCSC Genome Bioinformatics** (www.genome.ucsc.edu)
  March 2006 assembly Human Genome Brwoser
- **Ensembl Genome Browser** (www.ensembl.org)

2.6 List of reagents

The following table provides a list of the chemical composition of reagents produced in-house and used in the experimental techniques described before.

.

Reagents	Application	Chemical composition
Running Buffer	Capillary electrophoresis	30% methanol (Fisher Scientific), 0.5% formic acid (Sigma) in HPLC dH_2O (Fisher Scientific)
Standard Reconstitution Buffer	Commercial ELISA of Aβ	55 mM Sodium Bicarbonate Buffer (Sigma), pH 9.0
Orange G® Loading Buffer	DNA gel electrophoresis	0.06% Orange G® (Sigma), 50% glycerol (BDH) and 50% dH_2O (Millipore)
20x TBE Buffer	DNA gel electrophoresis	216 g Tris base (Sigma), 110 g boric acid (BDH), 18.6 g EDTA (BDH), made to 1 litre with dH_2O (Millipore)
Coating Buffer	ELISA of Aβ	0.1 M $NaHCO_3$ (Sigma), 0.1 M Na_2CO_3 (Sigma) pH 9.6 in dH_2O (Millipore)
Block Ace® Solution	ELISA of Aβ	1% Block Ace® (Serotec Ltd), 0.05% NaN_3 (Sigma), 5% of 20x PBS in dH_2O (Millipore)
Buffer EC*	ELISA of Aβ	0.02 M NaH_2PO_4 (Sigma), 2 mM EDTA (BDH), 0.4 M NaCl (Sigma), 0.2% BSA (Sigma), 0.05% CHAPS (Sigma), 0.4% Block Ace® (Serotec Ltd), 0.05% NaN_3 (Sigma), pH 7.0
Buffer C*	ELISA of Aβ	0.02 M NaH_2PO_4 (Sigma), 2 mM EDTA (BDH), 0.4 M NaCl (Sigma), 1.0% BSA (Sigma), 0.002% thimerosol (Sigma), pH 7.0
Coating Buffer	ELISA of CFH	0.1 M Glycine (Sigma) pH 9.5

Reagents	Application	Chemical composition
Wash Buffer	ELISA of CFH	PBS containing 0.05% Tween® (Sigma)
Sample Diluent	ELISA of CFH	PBS containing 0.1% BSA (98% lyophilised powder, Sigma) and 0.05% Tween® (Sigma)
Stop Solution	ELISA of CFH	1N Sulphuric acid (Sigma)
Blocking Solution	ELISA of CFH	1% BSA (98% lyophilised powder, Sigma) in dH_2O (Millipore)
20x PBS	General	0.4% KCl (BDH), 0.4% KH_2PO_4 (BDH), 2.3% Na_2HPO_4 (BDH), 16.2% NaCl (BDH) pH 7.4 in dH_2O (Millipore)
PBS-Tween®	General	PBS containing 0.05% Tween® (Sigma)
6x SDS Loading Buffer	Protein separation by Laemmli	35% 1M Tris HCl (BDH) pH 6.8, 30% glycerol (Sigma), 1% SDS (BDH), 0.9% DTT (Sigma), 0.01% bromophenol blue (BDH) in dH_2O (Millipore)
5x SDS Running Buffer	Protein separation by Laemmli	1.5% Tris-base (Sigma), 7.2% Glycine (BDH), 25% SDS in dH_2O (Millipore)
Transfer Buffer	Semi-dry transfer of proteins to membranes	0.3% Tris-base (Sigma), 1.4% Glycine (BDH), 20% methanol (Fisher Scientific)

Reagents	Application	Chemical composition
Fixing Solution	Silver staining	30% of 100% ethanol (Fisher Scientific), 10% of 100% glacial acetic acid* (Fisher Scientitic), 60% dH_2O (Millipore)
Staining Solution	Silver staining	0.1% $AgNO_3$ solution** (Sigma) in dH_2O from 20% stock.

* Buffer EC and Buffer C, were incubated at 56^{o}C in a waterbath for 40 minutes, cooled to room temperature and stored at 4^{o}C.
** Handled in the fume hood.

CHAPTER 3

ANALYSIS OF POWER

3.1 Introduction

Large-scale association studies, which aim to identify genetic variants that predispose individuals to common, complex diseases, are becoming a promising tool for deciphering the genetic basis of complex diseases. It is therefore important that the most appropriate sample size and a suitable genotyping platform are selected that best suit the population of interest.

Power is defined as the probability that a direct or indirect test of association between a genetic marker and the phenotype will be statistically significant given that such an association exists. Power depends on a number of factors, including the magnitude of genetic effect, sample size, prior probability and the type I error rate, α. In association studies, assumptions are made about the effect size of the genetic variant and the level of linkage disequilibrium (LD) between the marker and the causative variant as well as their respective allele frequencies.

In this chapter, the results of an empirical and predictive analysis of power, using the association between *APOE* genotypes and concentrations of plasma total cholesterol (TC), triglyceride (TG), low-density lipoprotein (LDL) and high-density lipoprotein (HDL) cholesterol are presented. The aim was firstly to try and replicate the well-established relationship between *APOE* genotypes and levels of circulating TC, TG, LDL and HDL and secondly, to estimate the power to detect variants of small to large effect size, using a predictive approach in the Vis sample set.

APOE protein variants have been shown to affect levels of triglyceride and very low density lipoproteins (vLDL) in human plasma (Utermann *et al*., 1979). Variation in human APOE protein, as a result of three common polymorphisms, have been reported to influence plasma cholesterol and triglyceride, and to confer susceptibility to type III hyperlipoproteinemia (Schneider *et al*., 1981, Breslow *et al.*, 1982, Rall *et al*., 1982). The APOE isoforms interact differently with specific lipoprotein receptors, which can alter circulating levels of cholesterol (Dong *et al*., 1994). Studies have associated *APOE*E2* carriers with lower levels of plasma total and LDL cholesterol than in *APOE*E3* homozygous individuals (Schneider *et al*., 1981; Weisgraber *et al*., 1982; Utermann *et al*., 1985; Sing & Davignon, 1985). The *APOE*E4* allele displays an opposite pattern, with higher levels of total and LDL plasma cholesterol (Utermann *et al*., 1985; Sing & Davignon, 1985).

3.2 Results

3.2.1 *APOE* genotyping in Vis sample set

Genomic DNA from 584 individuals from the village of Komiza, and 457 individuals from the village of Vis, on the Croatian island of Vis, were amplified using PCR. The *APOE* genotype of each sample was established after overnight digestion with *Cfo*I and running on an agarose gel.

The software PedCheck (http://watson.hgen.pitt.edu, University of Pittsburgh) was used to detect any genotyping incompatibilities based on the pedigree data already assembled. Furthermore, *APOE* allelic and genotypic frequencies were determined in both villages of Komiza and Vis separately, using the software SPSS v12.0 (**Table 3.1**). Frequencies were based on 555 individuals from Komiza (29 missing) and 437 from Vis (20 missing), either due to missing DNA samples or lack of genotypic data after repeated digestion and sequencing experiments.

As Table 3.1 shows, *APOE* allele frequencies observed in our population sample from Croatia agree with previous findings in Mediterranean populations with a long established agricultural economy, having a high frequency of the *APOE*E3* allele and relatively low frequencies of the *APOE*E4* and *APOE*E2* alleles. Comparison of allele and genotypic frequencies in Vis with published data from unrelated individuals in the U.K. population (Corbo *et al*., 1995; Lucotte *et al*., 1997), revealed a reduction in the *APOE*E4* allele frequency from 14% in the U.K. to 8% in the village of Komiza. The frequency of the homozygous *APOE*E4* genotype was also found to be reduced from 1.6% in the U.K. general population to 0.5% in the village of Komiza. An increase in the *APOE*E2* homozygote frequency of 0.3% in the U.K. to 4% in Komiza was also observed. Although the *APOE*E2* and *APOE*E4* alleles are far less frequent than the *APOE*E3* allele, and their homozygous genotypes are rare, the sample set was found to be consistent with Hardy-Weinberg equilibrium ($\chi^2 = 1.62$, $P = 0.899$).

	*APOE a*llele frequency (%)			*APOE g*enotype frequency (%)					
	E2	*E3*	*E4*	*E2/E2*	*E2/E3*	*E2/E4*	*E3/E3*	*E3/E4*	*E4/E4*
Komiza	8.0	84.0	8.0	4.0	14.8	1.6	70.1	12.6	0.5
Vis	6.0	83.0	11.0	1.1	10.1	0.2	68.4	19.2	0.9
U.K.	8.0	78.0	14.0	0.3	14.3	2.0	58.9	23.0	1.6

Table 3.1 *APOE* Allele and genotype frequencies in Vis.

APOE allele and genotype frequencies were measured in sample sets from the villages of Komiza (*N* = 555) and Vis (*N* = 437) compared with published data on the U.K. population (Corbo *et al*., 1995; Lucotte *et al*., 1997). All genotype frequencies were in Hardy-Weinberg equilibrium.

3.2.2 Empirical analysis of power using plasma lipids

An empirical analysis of power was carried by trying to replicate the association between *APOE* genotypes and plasma lipid concentrations in Vis. The complete set of *APOE* data from the island of Vis was used to evaluate the relationship between an individual's *APOE* status and normalised plasma concentrations of HDL, LDL, TC, and TG in a univariate analysis, using the statistical software package SOLAR (Almasy and Blangero, 1998).

The narrow-sense heritability (h^2) defined as the ratio of a trait's additive variance to its total variance, is a measure of the predictability of offspring trait values based on parental trait values.

In the present study, narrow-sense heritabilities were estimated using a polygenic model which uses information on the covariance among relatives to estimate the proportion of the trait variance that is due to additive genetic factors, environmental factors and covariates. The relatedness of individuals was taken into account by assigning the related individuals a shared environmental grouping. This consisted of the mother's ID in each pedigree. *APOE* genotypes were used as covariates in the univariate analysis of plasma lipids based on the *APOE* allelic and genotypic models.

The normality of the plasma lipid distributions were checked prior to the univariate analysis of *APOE* genotypes and plasma lipid concentrations, using the Kolmogorov-Smirnov test for normality. Deviations from normality were checked using the statistical software SPSS (v.12). Distributions of plasma HDL, LDL, TC and TG concentrations were all found to deviate significantly from normality (P_{HDL} = 4.5E-34; P_{LDL} = 1.1E-12; P_{TC} = 1.9E-09; P_{TG} = 1.0E-85). This emphasised the need to transform the data to give normal distributions in order to reduce the probability of false-positive results in future analyses. Plasma HDL, LDL, TC and TG measurements were normalised successfully using square, square-root, natural log and natural log transformations respectively ($P_{Kolmogorov\text{-}Smirnov} < 0.01$) using the software package SPSS v.12.

Narrow-sense heritability of plasma HDL concentration was estimated in the Vis sample set using a square transformation and was found to be non-significant at 0.133 (± 0.10 SE, P = 0.203) using sex, age and BMI as significant covariates (**Table**

3.2). The *APOE*E2* allele was significantly associated with an increase in plasma HDL ($\beta = 0.22 \pm 0.09$; $P_{add} = 0.026$, $P_{dom} = 0.014$) explaining 3.5% of the phenotypic variance ($R^2 = 0.035$). No significant associations were detected between *APOE*E3*, *APOE*E4* and plasma HDL in this sample set.

Square-root transformed plasma LDL concentration showed a narrow-sense heritability of 0.17 (± 0.11 SE, $P = 0.042$) with sex and age as significant covariates (**Table 3.2**). The most significant associations with LDL were observed for *APOE*E2* and *APOE*E4* alleles using a dominant model ($P_{APOE2} = 2.2E-09$, $P_{APOE4} = 1.6E-05$), explaining 11.2% and 10.1% of the trait variance respectively (**Table 3.2**). Furthermore, the *APOE*E2* allele was associated with a decrease in LDL levels ($\beta = -0.530$). The effect associated with *APOE*E4* was in the opposite direction ($\beta = 0.364$).

Analysis of plasma LDL using the genotypic model showed highly significant effects associated with the *E2*/*E3* ($P = 1.7E-08$, $\beta = -0.538$) genotypes and *E3*/*E4* ($P = 5.4E-07$, $\beta = 0.441$) genotypes, in support of the dominant model fitted to the *E2* and *E4* alleles (**Table 3.2**). However, no association was observed between *E2* or *E4* homozygous genotypes and plasma LDL. This is discussed in the summary and conclusions section of this chapter.

Measurement of natural logarithm-transformed plasma total cholesterol (TC) showed a non-significant narrow-sense heritability of 0.145 (± 0.08 SE, $P = 0.093$), using sex, age, age-by-sex interaction and BMI as covariates (**Table 3.2**). Highly significant evidence for dominant effects on TC were associated with *APOE*E2* ($P = 4.0E-08$, $\beta = -0.504$) and *APOE*E4* ($P = 1.2E-05$, $\beta = 0.331$) alleles. The *APOE*E2* and *APOE*E4* alleles were found to explain 15.1% and 14.2% of the TC trait variance respectively. The effect associated with the *APOE*E3* allele on plasma TC concentration did not differ significantly from zero. In the genotypic model, a highly significant association was observed between the *E2*/*E3* genotype and TC ($P = 1.2E-07$, $\beta = -0.528$) and between *E3*/*E4* genotype and TC ($P = 2.0E-06$, $\beta = 0.403$). Each of the genotypes explained about 14.5% of the phenotypic variance. No significant association was observed between *APOE* genotypes and the concentration of plasma triglyceride.

Trait (mean±SD) (Transformation)	h^2 (S.E.) (P-value)	*APOE* allele/genotype	*P* (Allelic model) *add.*	*dom.*	*P* (Genotypic model)	β (S.E.)	R^2
HDL	0.133 (0.10)						
(3.9±1.03)	(0.203)						
(square)		*E2*	**0.026**	**0.014**		0.220 (0.09)	0.035
		E3	0.608	0.556		-	0.026
		E4	0.165	0.172		-	0.026
		E2/E2			0.989	-	0.026
		E2/E3			0.054	-	0.035
		E2/E4			0.132	-	0.031
		E3/E3			n.a.*	-	
		E3/E4			0.182	-	0.031
		E4/E4			0.632	-	0.031
LDL	0.170 (0.11)						
(7.1±1.04)	(0.042)						
(square root)		*E2*	**1.9E-08**	**2.2E-09**		-0.530 (0.08)	0.112
		E3	0.239	0.096		-	0.088
		E4	**7.8E-05**	**1.6E-05**		0.364 (0.08)	0.101
		E2/E2			0.346	-	0.089
		E2/E3			**1.7E-08**	-0.538 (0.09)	0.112
		E2/E4			0.157	-	0.089
		E3/E3			n.a.*	-	
		E3/E4			**5.4E-07**	0.441 (0.09)	0.107
		E4/E4			0.675	-	0.089
TC	0.145 (0.08)						
(9.2±1.09)	(0.093)						
(natural log)		*E2*	**1.1E-07**	**4.0E-08**		-0.504 (0.09)	0.151
		E3	0.373	0.207		-	0.129
		E4	**5.1E-05**	**1.2E-05**		0.331 (0.08)	0.142
		E2/E2			0.156	-	0.129
		E2/E3			**1.2E-07**	-0.528 (0.09)	0.146
		E2/E4			0.703	-	0.129
		E3/E3			n.a.*	-	
		E3/E4			**2.0E-06**	0.403 (0.09)	0.144
		E4/E4			0.783	-	0.129
TG	0.270 (0.14)						
(0.4±0.47)	(0.027)						
(natural log)		*E2*	0.734	0.488		-	0.139
		E3	0.220	0.455		-	0.137
		E4	0.201	0.247		-	0.139
		E2/E2			0.293	-	0.139
		E2/E3			0.537	-	0.139
		E2/E4			0.137	-	0.139
		E3/E3			n.a.*	-	
		E3/E4			0.571	-	0.139
		E4/E4			0.370	-	0.139

Table 3.2 Univariate analysis of *APOE* genotypes and plasma lipids concentrations.

1,030 unselected samples, consisting of related and unrelated individuals, were used in the univariate analysis *APOE* genotypes and levels of plasma HDL (high density lipoprotein), LDL (low density lipoprotein), TC (total cholesterol) and TG (triglyceride). β is the regression coefficient which represents the change in mean plasma concentration of each normalised trait distribution, using allelic (additive and dominance) and genotypic models of *APOE* as covariates. R^2 shows the proportion of the phenotypic variance explained by each covariate. Narrow-sense heritabilities (h^2) were estimated for each trait. Statistically significant results are shown in bold. n.a., not applicable. All analyses were carried out using the software SOLAR (Almasy & Blangero, 1998).

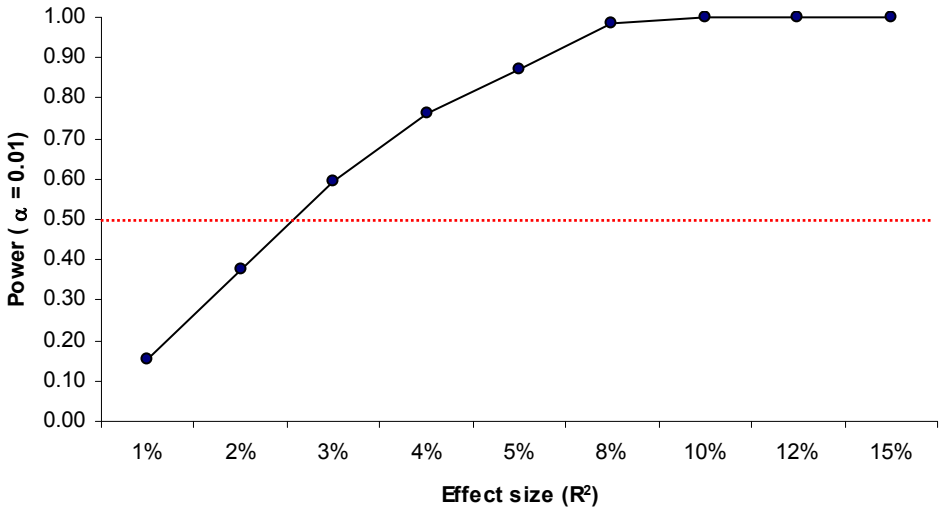
The narrow-sense heritability of TG after natural logarithm-transformation was estimated to be 0.27 (± 0.14 SE, P = 0.027) (**Table 3.2**). Sex, age, BMI and smoking were all found to be significant covariates for changes in plasma TG concentration and were included in the regression model.

The effect of the *APOE* genotype on plasma lipid concentrations, which is an example of a large effect genetic influence, was therefore supported in samples from Vis. The results of the empirical analysis of power showed that there was sufficient power available to detect genetic associations of effect size (R^2) greater than 3% in the Vis data set (**Table 3.2**).

3.2.3 Predictive analysis of power

The power to detect a quantitative trait locus (QTL) influencing a quantitative trait by genetic association was estimated in a sample of 500 unrelated individuals. Analysis was carried out using the programme Genetic Power Calculator (Purcell *et al*., 2003). Assuming a 1% threshold for a type I error rate, power was estimated for QTL of varying effect size and frequency. The proportion of the phenotypic variance explained by the QTL (R^2) was taken as representative of the effect size of that QTL.

The results showed that there was insufficient power in 500 unrelated individuals to detect a QTL with an effect size of 1% (Power = 0.15) (**Figure 3.1**). Using the same number of unrelated individuals, the power to detect a QTL of 2%, 3%, 4% and 5% effect size increased to 0.38, 0.60, 0.76 and 0.87, respectively with marker allele frequency of 0.05 - 0.07 (**Figure 3.1**). Results showed an exponential increase in the available power (0.15 to 0.99) as the QTL effect size increased from 1% to approximately 8% (**Figure 3.1**). The power to detect QTL of large effect size (10% and 15%) was predicted to increase exponentially to 0.998 (R^2 = 10%) and 1 (R^2 = 15%) (**Figure 3.1**).



Marker ID	Marker allele frequency	R^2	Power (α = 0.01)	Effect size
M1	0.05	0.01	0.152	1%
M2	0.05	0.02	0.377	2%
M3	0.05	0.03	0.596	3%
M4	0.05	0.04	0.764	4%
M5	0.07	0.05	0.870	5%
M6	0.10	0.08	0.987	8%
M7	0.17	0.10	0.998	10%
M8	0.20	0.12	0.999	12%
M9	0.20	0.15	1.000	15%

Figure 3.1 Predicted power for detecting polymorphisms of small to large effect size, influencing a quantitative trait by genetic association.

Predictive power for the detection of a genetic locus influencing a quantitative trait in 500 unrelated individuals shows that there is sufficient power (≥ 0.50) available in the Vis dataset to detect polymorphisms of $\geq$ 3% effect size. Power was estimated for markers of small, moderate and large effect size as measured by the proportion of phenotypic variance explained by the marker (R^2). The marker and the causative variant are assumed to be in strong linkage disequilibrium (> 0.80) and to have the same allele frequencies. α is the type I error rate. Data was obtained using the Genetic Power Calculator® (Purcell *et al*., 2003)

3.2.4 Summary and discussion

In this chapter, the relationship between *APOE* genetic variants and concentrations of plasma HDL, LDL, TC and TG was used to evaluate power empirically by attempting to replicate genetic effects associated with *APOE* locus variants on plasma lipid concentrations. Power was also estimated for genetic variants of small to large effect using the variance components methodology implemented in the Genetic Power Calculator® (Purcell *et al*., 2003).

Firstly, the *APOE* genotypes of 1,021 individuals from Vis island were determined, using PCR and RFLP analysis, in order to carry out a univariate analysis of the association between *APOE* genotypes and plasma concentrations of HDL, LDL, TC and TG. The regression model also included smoking, BMI, age and sex as covariates. Lower plasma HDL concentrations have been reported in males and smokers compared with females and non-smokers (Bihari-Varga *et al*., 1981). Smoking has also been shown to be associated with elevated plasma LDL and TG concentrations in disease-free individuals (Freeman *et al*., 1993). Plasma concentrations of TC and TG have been reported to increase with increasing age and BMI in disease-free individuals (Mann *et al*., 1988). Furthermore, male smokers affected with coronary artery disease (CAD), for which plasma lipids are strong risk factors, have been shown to have reduced concentrations of plasma HDL compared with female non-smokers (Hill *et al*., 1991). Increased concentrations of plasma TG have been reported in females affected with CAD (Hill *et al*., 1991).

The significant covariates described above were used to construct a polygenic model in order to estimate narrow-sense heritabilities (h^2) of plasma lipid concentrations. However, significant estimates of heritability were only observed for plasma LDL ($h^2 = 0.17$, $P = 0.042$) and TG ($h^2 = 0.27$, $P = 0.027$) (**Table 3.2**). A low or zero additive variance places a constraint on the magnitude of additive effects of a QTL, however a high additive variance does not necessarily indicate that any QTL follows a strictly additive model, since it is possible to have a high additive genetic variance even when all loci follow a dominant model (Abney *et al*., 2001). It is also possible to have a trait that is strongly influenced by genetic variation that has a relatively low additive variance (i.e. low h^2), for example if selection has been acting on the trait (Fisher, 1958; Charlesworth & Hughes, 2000). A high heritability

estimate would however increase the power for detecting common susceptibility variants with an additive effect.

Published data have reported significant heritabilities for plasma lipid concentrations as summarised in **Table 3.3**. Heritability estimates of plasma LDL ($h^2 = 0.17$, $P < 0.05$) and TC ($h^2 = 0.28$, $P < 0.05$) in a sample of 868 individuals from the isolated Rucphen (Dutch) population were found to be similar to those observed in the Vis population (Isaacs *et al*., 2007) (**Table 3.3**). However, in the Rucphen population, significant heritabilities were found for plasma HDL ($h^2 = 0.51$, $P < 0.05$) and TC ($h^2 = 0.19$, $P < 0.05$) in contrast to the findings in the Vis data set (**Table 3.3**). This could be due to the lack of power in the Vis data set (472 related individuals) compared with the Rucphen samples (868 related individuals) due to the lower number of related individuals from whom the heritability estimates are extracted. Furthermore, distant relationships within the Vis pedigrees (despite the presence of a single large pedigree of 125 related individuals) could account for the reduced power.

The large difference between the heritabilities observed for HDL in Vis and Rucphen samples could be the result of extreme isolation and increased inbreeding in Vis compared with the Rucphen community (Isaacs *et al*., 2007) since higher estimates of heritability in less inbred populations could be due to increased variation in the trait distribution. Inbreeding can also result in diversity by creating a higher proportion of extremes in the population (Rudan & Campbell, 2004). Indeed, lower variation in HDL distribution has been reported in the Vis population compared with the urban Croatian population (Polasek *et al*., 2006).

Compared with Vis, significantly higher heritabilities have been reported for plasma lipids in 806 individuals from the founder population of Hutterites (Abney *et al*., 2001), in a sample of 160 pairs of unselected Dutch twins (Beekman *et al*., 2002) and in an urban population series of 85 Caucasian families ascertained through a hyperlipidemic proband (Edwards *et al*., 1999) (**Table 3.3**).

Trait	Vis (*n* =1,030)	Hutterites (*n* =806)	Rucphen (*n* =868)	Dutch twins (*n* =320)	GET Study, USA (*n* =85 families)
HDL	0.13	0.63*	0.51*	0.75*	0.54*
LDL	0.17*	0.50*	0.17*	0.83*	0.34*
TC	0.15	-	0.19*	0.82*	-
TG	0.27*	0.37*	0.28*	0.71*	0.41*

Table 3.3 Comparison between reported estimates of heritability of plasma lipid concentrations and estimates obtained in the Vis population.

Narrow-sense heritability estimates of plasma lipid concentrations were found to vary between the Vis population and published data on related individuals from the founder population of Hutterites (Abney *et al*., 2001), related individuals from the isolated Dutch (Rucphen) population (Isaacs *et al*., 2007), 160 pairs of unselected Dutch twins (Beekman *et al*., 2002) and Caucasian families identified through a hyperlipidemic proband in the Genetic Epidemiology of Hypertriglyceridemia Study (GET) (Edwards *et al*., 1999). HDL, high-density lipoprotein; LDL, low-density lipoprotein; TC, total cholesterol; TG, triglyceride. Asterisks denote statistically significant results ($P < 0.05$).

		Vis, Croatia					
		Effect of the *APOE* allele (β)			R^2 (%)		
		E2	***E3***	***E4***	***E2***	***E3***	***E4***
	Allele frequency	0.079	0.830	0.091			
HDL		0.22	n.s.	n.s.	3.5	n.s.	n.s.
LDL		-0.53	n.s.	0.36	11.2	8.8	10.1
TC		-0.50	n.s.	0.33	15.1	12.9	14.2
TG		n.s.	n.s.	n.s.	n.s.	n.s.	n.s.

		Ottawa, Canada (Sing & Davignon, 1985)			
		Effect of the *APOE* allele			R^2 (*APOE* locus, %)
		E2	***E3***	***E4***	
	Allele frequency	0.078	0.770	0.152	
HDL		-1.44	-0.03	0.90	n.s.
LDL		-13.06	0.01	6.68	8.3
TC		-12.60	-0.29	7.99	6.9
TG		3.94	0.36	-3.86	1.0

Table 3.4 Comparison of *APOE* association with plasma lipid concentrations between Vis and published data.

APOE allele frequencies and the allele-specific effects (β) on plasma lipid concentrations obtained from 1,030 unselected individuals from Vis compared with published data (Sing & Davignon, 1985) on the direction of the *APOE* allele-specific effects on plasma LDL and TC concentrations. The proportion of the phenotypic variance explained by *APOE* allelic variants (R^2) was found to be higher in the Vis population compared with the randomly selected cohort of 122 unrelated individuals from Ottawa (Sing & Davignon, 1985). HDL, high-density lipoprotein; LDL, low-density lipoprotein; TC, total cholesterol; TG, triglyceride. N.S., non-significant

Inconsistencies in the estimated h^2 between populations could also be the result of differences in sample sets used in different studies (**Table 3.3**) as the number of related individuals (the degree of relatedness in the data set) and the type and depth of the pedigrees can affect the magnitude of h^2. More distant relationships may result in reduced power in estimating heritabilities (Hsu *et al*., 2005). The effect of the degree of relatedness is demonstrated by the significantly higher h^2 estimations in the Dutch twins (Beekman *et al*., 2002) compared with the general population samples (**Table 3.3**). However, heritability based on sibling pairs (especially twins) may include non-additive genetic effects and shared environment which contribute to greater phenotypic resemblance among sibling pairs than parent-offspring pairs (Hsu *et al*., 2005). Inclusion of parental data provides additional heritability through the parent-offspring covariance which is an unbiased estimate of additive genetic variance (Hsu *et al*., 2005). This may be the case in studies (**Table 3.3**) in which higher heritabilities are observed for some traits. The lack of significant h^2 estimates obtained in the Vis population can also be explained by the possible action of natural selection on these traits (Fisher, 1958; Charlesworth & Hughes, 2000) associated with different environmental exposures such as diet, smoking and exercise (Vuletić Mavrinac & Mujkić, 2006) which influence plasma lipid concentrations.

Analysis of plasma HDL showed suggestive association with the *APOE*E2* allele, under both dominant (P = 0.014) and additive (P = 0.026) models. Weak association is probably due to the low frequency of the *E2* allele in this population. No association was detected between other *APOE* alleles, or any of the *APOE* genotypes, and plasma HDL concentration. The lack of association between *APOE* and plasma HDL has also been reported in the literature (Sing & Davignon, 1985) (**Table 3.4**) suggesting that despite a high reported heritability (Abney *et al*., 2001; Isaacs *et al*., 2007; Beekman *et al*., 2002; Edwards *et al*., 1999) the variation in human plasma HDL concentration is determined by genetic factors other than *APOE*.

The results showed that reduced concentrations of plasma LDL (β = -0.530) are highly significantly associated with the *APOE*E2* allele (P = 2.2E-09), while raised plasma LDL concentrations (β = 0.364) are strongly associated with the *E4* allele (P = 1.6E-05). Under the allelic model, evidence of dominance was also observed for the *E2* and *E4* alleles. *APOE* genotypes, notably the *E2*/*E3* and *E4*/*E4*

genotypes, also showed significant associations with differences in plasma LDL concentrations ($P_{E2/E3}$ = 1.7E-08, $P_{E3/E4}$ = 5.4E-07). The lack of significant association between the *E2*/*E2* and *E4*/*E4* genotypes and plasma LDL is probably due to the low frequencies of those genotypes in the Vis sample set ($frequency_{E2/E2}$ = 0.011, $frequency_{E4/E4}$ = 0.007). The *APOE* association with plasma LDL is consistent with other reported data showing that variation at the *APOE* locus explains 8.3% of the total LDL variance (Sing & Davignon, 1985), compared with 11.2% for the *APOE*E2* and 10.1% for the *APOE*E4* alleles in Vis (**Table 3.4**). The stronger associations reported in Vis could be due to the larger sample size of 1,031 individuals from Vis compared with only 122 individuals from Ottawa (Sing & Davignon, 1985). More accurate family structure and trait measurements can also result in better estimates of the proportion of phenotypic variance explained (Dawn & Barrett, 2005).

Analysis of plasma TC showed highly significant evidence for a dominance effect associated with *APOE*E2* (P = 4.0E-08, β = -0.504) and *APOE*E4* (P = 1.2E-05, β = 0.331) alleles, explaining 15.1% and 14.2% of the total TC variance respectively. The effect on differences in plasma TC concentration associated with the *APOE*E3* allele did not differ significantly from zero. Analysis of TC and the *APOE* genotypic model showed highly significant associations between the *E2*/*E3* genotype and TC (P = 1.2E-07, β = -0.528) and between the *E3*/*E4* genotype and TC (P = 2.0E-06, β = 0.403), both explaining about 14.5% of the phenotypic variance. Despite the same direction of effect reported for the *APOE*E2* and *APOE*E4* alleles on plasma TC, the proportion of the phenotypic variance explained by those alleles in other studies has been considerably lower (R^2 = 6.9%) (Sing & Davignon, 1985) (**Table 3.4**). This could again be the result of larger sample size in Vis (N = 1,021) compared with the sample size used by Sing and Davignon (N = 122).

Finally, no significant association was observed between *APOE* genotypes and plasma triglyceride concentration despite obtaining a significant heritability estimate for this trait in Vis (**Table 3.2**). Variation at the *APOE* locus has been shown to explain 1% of the variance in plasma TG concentration in unrelated samples from Ottawa (Sing & Davignon, 1985) implying that variation in this trait in Vis may be explained by genetic factors other than *APOE* (**Table 3.4**).

A wider range of plasma TG and HDL concentrations and a smaller range of plasma LDL concentration was observed in the general Croatian population compared with the Vis population (Polasek *et al*., 2006). Although reduced genetic and environmental diversity in isolated human populations are expected to reduce the variance in observed phenotypic values (Rudan, 1999a; Rudan *et al*., 1999b), it appears that specific population genetic processes in this isolate, such as inbreeding and population substructure could be acting to maintain the variation in these traits (Rudan & Campbell, 2004).

In conclusion, the empirical analysis of power using the association between *APOE* genotypes and plasma lipid concentrations showed that there was sufficient power to replicate associations between *APOE* alleles and plasma LDL and TC concentrations with effect sizes ranging between 10-15%. Significant (but weaker) association was also observed between *APOE*E2* and plasma HDL (R^2 = 0.035, P = 0.01-0.03) showing that there is sufficient power to detect effect sizes low as 3.5%. Lack of power to detect an association of small effect size (1%), as reported for *APOE* and TG by Sing and Davignon (1985), was apparent in the Vis data set by the absence of association between *APOE* and TG.

Predictive analysis of power showed that the power to detect a causal variant increases with the effect size of the marker of interest (assumed to be in high LD with the causal variant) and with the marker allele frequency. The results of the predictive analysis of power, using the Genetic Power Calculator (Purcell *et al*., 2003), showed that in a QTL association analysis, sufficient power (≥ 60%) would be available for detecting variants with effect size of approximately 3% using 500 unrelated individuals (**Figure 3.1**). However, a power of only 15% and 37% would be available to detect variants of effect size 1% and 2%, respectively. The lack of power in detecting variants of small effect size (1-2%) is probably due to the small number of unrelated individuals in the Vis data set (N = 500) who contribute much of the power. Furthermore, the power may be underestimated due to higher extents of LD which has been show to exist in the Vis isolate compared with urban populations (Vitart *et al*., 2005; Vitart *et al*., 2006).

Klein (2007) estimated power for the Illumina Hap300 genotyping platform, as used in this study, and showed that it increases with increasing sample size and

magnitude of effect (**Figure 3.2**). Therefore, for a sample size of 500 unrelated individuals in a case/control study, there would be sufficient power (> 0.50) to detect variants with a genotype relative risk (GRR) greater than 2.5. GRR of 2.5 would be considered of moderate effect size when expressed in terms of the proportion of the phenotypic variance (R^2) in QT association analyses. In the same study, the number of unrelated individuals needed to obtain 80% power in detecting a polymorphism with a smaller effect size (GRR = 1.75) was estimated to be around 2,000 using the Illumina Hap300 array (light green bar in **Figure 3.2**). The number of unrelated individuals needed to detect a causative variant with a GRR of 1.75 and 80% power, was found to be higher when using the Affymetrix 550K array (dark green bar **Figure 3.2**).

These results reported by Klein were obtained from a CEPH population and emphasised that more power could be obtained by genotyping more individuals at fewer loci rather than fewer individuals at more SNPs.

In conclusion, both predicted and empirical power analyses were carried out, the latter using known biomarkers of dyslipidaemia. The effect of the *APOE* genotype on plasma lipid levels was replicated, confirming that adequate power is available. It is also predicted that there would be sufficient power in the Vis data set to detect polymorphisms with effect sizes as small as 3.0%.

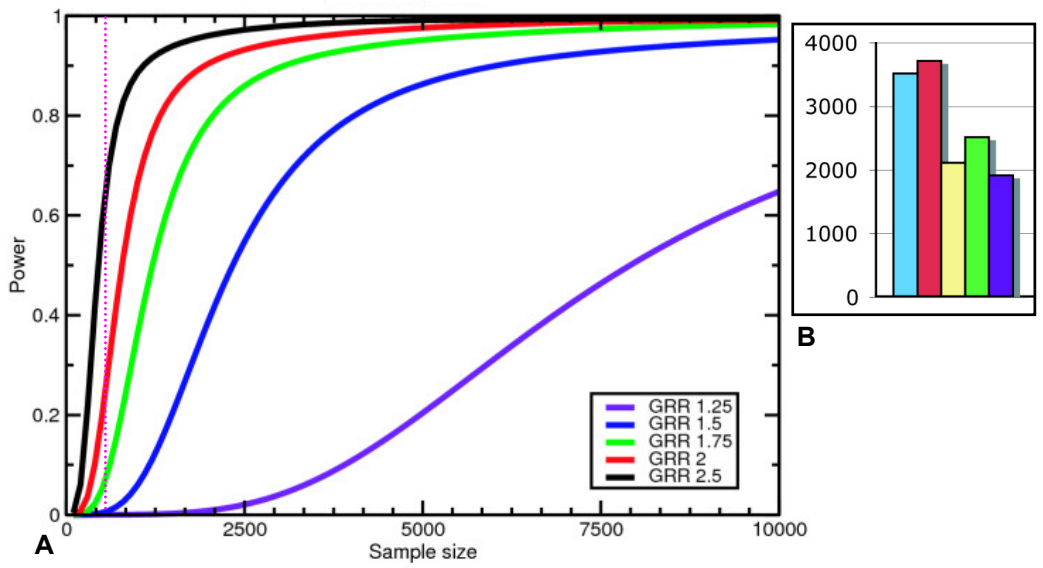


Figure 3.2 Analysis of power in a case-control genome-wide association study.

The graph shows the power estimated for varying genotype relative risks (GRR) for a test of genotypic association using the Illumina Hap300 array in a case-control cohort. **A.** For a sample size of 500 individuals, indicated by the dotted pink line, there would be sufficient power (> 0.50) to detect polymorphisms of GRR > 2.5. **B.** The bar chart shows the total number of individuals needed to achieve 80% power to detect a causative QTL assuming a GRR of 1.75. The light green bar represents the Illumina Hap300 array used in a CEPH population sample set. Blue, red, dark green and dark blue bars show the Affymetrix 250K *Nsp*, Affymetrix 250K *Sty*, Affymetrix 500K and Illumina 550K respectively. Diagram adapted from Klein, *BMC Genetics (*2007).

CHAPTER 4

BIOCHEMICAL AND LINKAGE ANALYSES OF PLASMA AMYLOID-β

4.1 Introduction

Brain amyloidosis, comprising extracellular deposits of amyloid-β (Aβ) peptides, and taupathy characterised by intracellular accumulation of hyperphosphorylated tau in the form of neurofibrillary tangles are the two main pathological features of the brains of patients with Alzheimer disease (Hardy & Selkoe, 2002). Amyloid-β [1-42] (Aβ42), tau and phosphorylated tau in cerebrospinal fluid (CSF) are established biomarkers for Alzheimer disease diagnosis (AD) (Kanai *et al*., 1998). Aβ42 aggregates more easily in the brain than the shorter amyloid-β [1-40] (Aβ40) due to the presence of two additional hydrophobic amino acids. Aβ42 is believed to be the initial species deposited in the brain, making it the key molecule in AD pathology (Hardy & Selkoe, 2002). Aβ is produced in the brain and cleared to the plasma via the CSF and the blood-brain barrier (Zlokovic *et al*., 2004). The mouse Apoe protein has been shown to participate in the Aβ transport through the blood-brain barrier (Bell *et al*., 2007).

It has been proposed that elevated concentrations of plasma Aβ42, the predominant form in senile plaques reflect a generalised increase in both neural and non-neural cells (Mayeux *et al*., 1999). It is well established that all of the mutations that cause early-onset familial Alzheimer disease (FAD) result in an increase in the plasma Aβ42 concentration (Scheuner *et al*., 1996). Furthermore, both plasma Aβ40 and Aβ42 concentrations are elevated in first-degree relatives of late-onset Alzheimer disease (LOAD) families (Ertekin-Taner *et al*., 2008). This study supports the notion that plasma Aβ concentrations show premorbid elevations in individuals who are genetically predisposed to LOAD. A prospective epidemiological study of 530 individuals ($N_{\text{AD cases}}$ = 165), in which the concentrations of plasma Aβ40 and Aβ42 were measured over three years, showed that compared with those who never developed AD, patients with AD at baseline and those who developed AD later had significantly higher plasma Aβ42 concentrations (Mayeux *et al*., 1999, 2003). These studies showed that in patients with newly acquired AD, the plasma Aβ42 concentration declined significantly compared with the control individuals or those with prevalent AD. The ratio of Aβ42 to Aβ40 (Aβ42:Aβ40) in plasma has previously been shown to be a good indicator of the

harmful levels of Aβ42 deposited in the brain (van Oijen *et al*., 2006; Graff-Radford *et al*., 2007). These studies showed that the risk of dementia was increased in individuals with high concentrations of plasma Aβ40 or low concentrations of plasma Aβ42 as a results of the deposition of Aβ42 in plaques prior to the onset of AD.

Narrow-sense heritabilities of both plasma Aβ40 and Aβ42 have previously been estimated to be around 41% and 56% respectively (Ertekin-Taner *et al*., 2001). Linkage analyses, using plasma Aβ as an intermediate phenotype, in extended LOAD families have identified peaks with significant (> 3.0) LOD scores on chromosome 10q (Ertekin-Taner *et al*., 2000; Ertekin-Taner *et al*., 2001). *APOE* variants are, to date, the only common genetic susceptibility factors which have been reliably associated with late-onset Alzheimer disease in multiple studies (Saunders *et al*., 1993; Corder *et al*., 1993). The *APOE*E4* allele has been associated with an increased risk of LOAD by lowering the age of onset in a dosage-dependent manor (Strittmatter *et al*., 1993; Basun *et al*., 1995).

The analysis of plasma Aβ in this study is based on the hypothesis that Aβ40 and Aβ42 peptides are produced and secreted by body cells (predominantly by the brain) as a result of genetic predisposition and environmental factors such as oxidative stress (Ross & Poirier, 2004). The fractions of Aβ40 and Aβ42 which are released into the plasma (resulting in an increase in plasma Aβ concentrations) reflect the CSF concentration, which is in turn influenced by the severity of aggregation in the brain (Mayeux *et al*., 1999, 2003) and both increase with age. However, a reduction in the plasma concentration of Aβ42 about the time of onset of Alzheimer disease (shown by a reduction in the ratio of Aβ42:Aβ40) is thought to represent an increase in the rate of deposition of Aβ42 in the brain (Hardy & Selkoe, 2002) based on the analysis of the Tg2576 mouse model of AD which over-produces Aβ42 (DeMattos *et al*., 2002). In this chapter, the results of the assay of human Aβ40 and Aβ42 in plasma are presented. The aim is to use plasma Aβ40 and Aβ42 as biomarkers for Alzheimer disease to identify genetic and environmental factors which influence them in a general population sample.

Plasma Aβ40 and Aβ42 concentrations were firstly measured by ELISA. Covariate effects, including *APOE* genotype, associated with circulating concentrations of Aβ40, Aβ42 and Aβ42:Aβ40 ratios were also assessed. Furthermore, a genome-wide linkage analysis using 810 microsatellite markers was carried out in order to identify QTL associated with differences in plasma Aβ40 and Aβ42 concentrations in a general population sample from the Croatian island of Vis.

A number of assays for the measurement of Aβ40 and Aβ42 are available commercially, however, their use in plasma has not been validated. One such assay, available from Biosource®, was initially tried but failed to produce reliable results for reasons that are explained later. Two "sandwich" ELISAs (Suzuki *et al*., 1994), developed in the laboratory of Professor S. G. Younkin (Mayo Clinic, Jacksonville, USA) and optimised for use with EDTA anticoagulated plasma, were employed to measure circulating concentrations of Aβ40 and Aβ42 in 999 fasting plasma samples collected from the island of Vis (**Figure 4.1**). The assay involved two mouse monoclonal capture antibodies, BAN50 (detecting Aβ1-16) and BNT77 (detecting Aβ11-28), which were used to coat ELISA microplates (Suzuki *et al*., 1994). Two mouse monoclonal antibodies, BA27 (detecting Aβ1-40) and BC05 (detecting Aβ35-43) were then used as detection antibodies for the specific detection of Aβ40 and Aβ42 respectively (Suzuki *et al*., 1994) (**Figure 4.1**).

4.2 Results

4.2.1 Commercial assay of Aβ40 and Aβ42 in plasma

Eighty-eight EDTA plasma samples from the village of Komiza on Vis island were used in the initial quantitation of the two Aβ species (Aβ40 and Aβ42) using a commercial sandwich ELISA kit (Biosource®). *In vitro* quantitative determination was carried out in duplicate for each sample (diluted 1 in 5) and each 96-well plate also contained serial dilutions of the standard peptide of known concentration and a blank control. Despite obtaining reliable standard curves (**Figure 4.2**), from which the concentration of each sample was deduced, the values obtained for individual plasma samples (in duplicate) were highly variable.

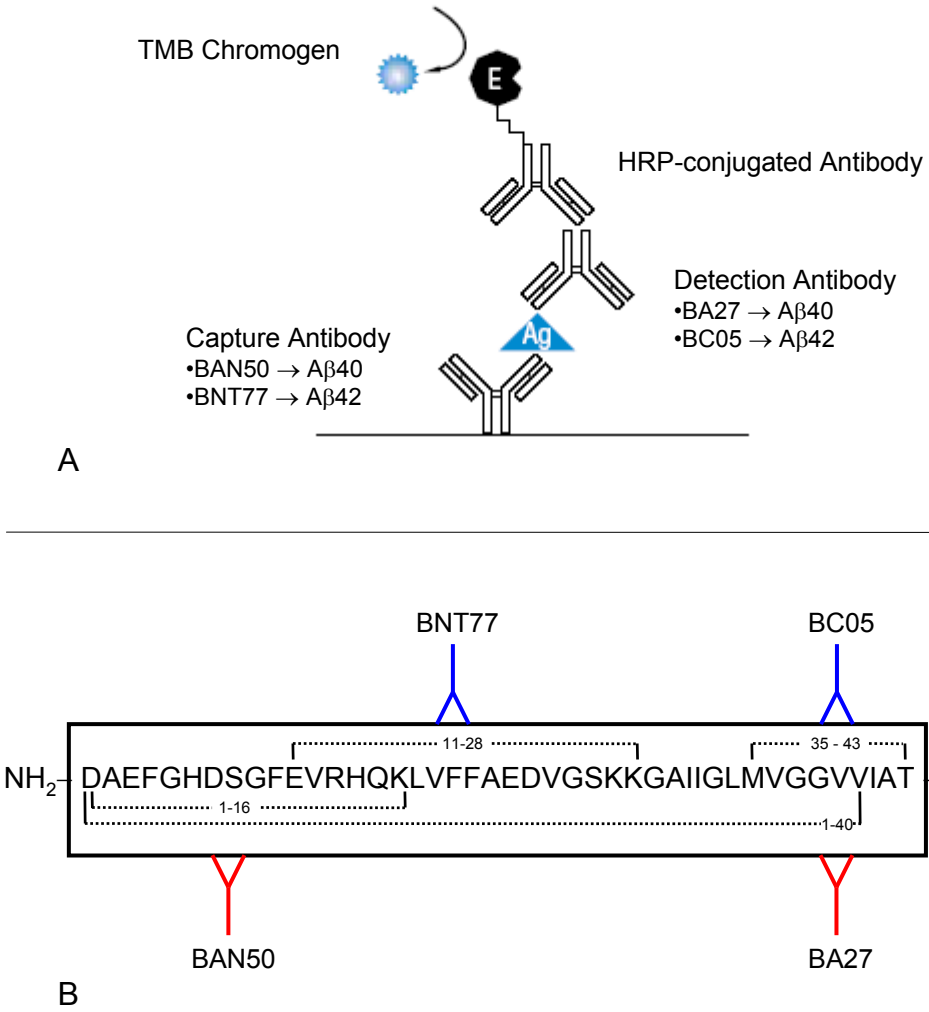
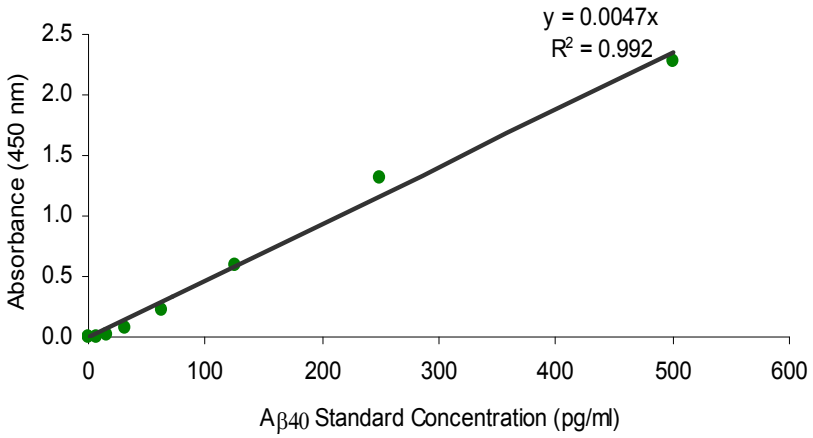
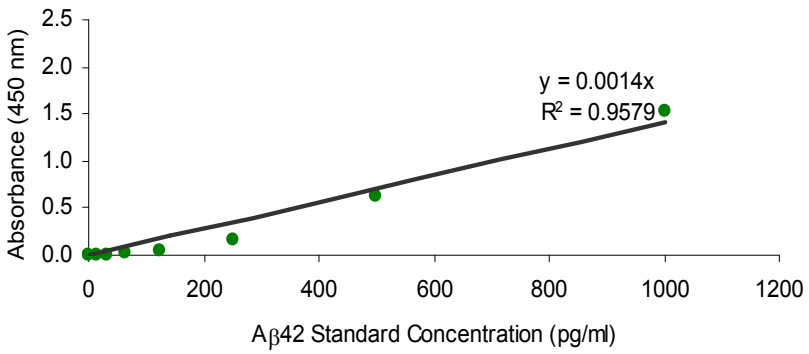


Figure 4.1 Diagrammatic representation of the "sandwich" ELISAs of plasma Aβ40 and Aβ42.

Monoclonal capture antibodies, BAN50 and BNT77 were paired with monoclonal detection antibodies BA27 and BC05 in two independent assays to detect plasma Aβ40 and Aβ42 concentrations by fluorometric detection at 450 nm **(A)**. The BAN50 antibody recognises residues 1-16 and the BNT77 antibody recognises residues 11-28. BA27 (epitope: Aβ1-40) and BC05 (epitope: Aβ35-43) specifically detect Aβ40 and Aβ42 peptides respectively, as shown on the amino acid sequence of the 43-residue Aβ peptide **(B)**. Details of antibody recognition sites were obtained from Suzuki *et al*., 1994.



A



B

Figure 4.2 Standard curves of the two commercial assays of Aβ40 and Aβ42 in plasma.

Standard curves obtained by serially diluting the standard peptide of known concentration for Aβ40 **(A)** and Aβ42 **(B)** assays, including a blank control. The equation of the best-fitted line was obtained by linear regression (Microsoft Excel) and used to measure the exact concentration of the two Aβ species in EDTA plasma samples. The squared correlation coefficient (R^2), for each line is also shown.

Binding of Aβ to carrier molecules results in the sequestration of the Aβ epitope and therefore failure to detect the bound portion of the peptide in plasma (Kuo *et al*., 2000). Therefore, pre-treatment of plasma samples using extremes of pH was suggested, which would be used to separate the bound fraction of Aβ from its carrier proteins. Samples were pre-treated with different dilutions of acid (formic acid 95%) and alkali (5M NaOH) solutions for 30 minutes and then neutralised. Each neutralised sample was then run on a 15% polyacrylamide gel, which included synthetic Aβ peptide as a positive control, and transferred to a PVDF membrane for probing with primary and secondary antibodies. However, no clear results were obtained, possibly because Aβ concentrations in plasma were below the threshold concentrations required for antibody detection. This confirmed the need for a more sensitive method of Aβ detection in human plasma.

4.2.2 Optimised assay of Aβ40 and Aβ42 in plasma

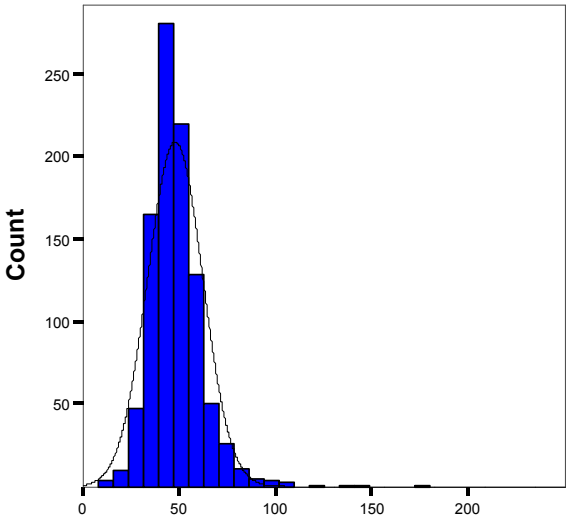
This work was carried out as part of a collaboration with Professor S. G. Younkin. A "sandwich" ELISA was used, which had been specifically designed and optimised for measurement of Aβ40 and Aβ42 in EDTA anticoagulated plasma (Suzuki *et al*., 1994). The problem of sequestration by carrier molecules in plasma was overcome by the antibodies having a higher affinity for Aβ peptides (Suzuki *et al*., 1994).

Quantitative determination of plasma Aβ40 and Aβ42 concentrations was carried out in triplicate using 999 plasma samples collected from the island of Vis. Previously unthawed EDTA plasma samples were used in the ELISA as both Aβ peptide species have been shown to be highly sensitive to freeze-thawing (S. G. Younkin, personal communication). In a second round of quantitation, sample replicates with high coefficients of variation (CV > 20%) were retested and a new set of measurements was obtained which was combined with the previous measurements. Outlier values within replicates were excluded using the Grubb's test for outlier detection assuming a normal distribution between replicated values. However, samples in which the three replicates showed a normal distribution but still had a high CV were excluded. For Aβ40 measurements, the CV threshold for exclusion was set at three standard deviations (SD) above the mean CV of 11.7 (±

11.7, SD). For Aβ42 measurements, the CV threshold was also set at three standard deviations above the mean CV of 24.1 (± 22.2, SD). Therefore, 50 individuals from the Aβ40 and 51 individuals from the Aβ42 distributions were excluded based on the above criteria. This preliminary investigation was carried out in order to minimise the probability of obtaining false positive results due to measurement error.

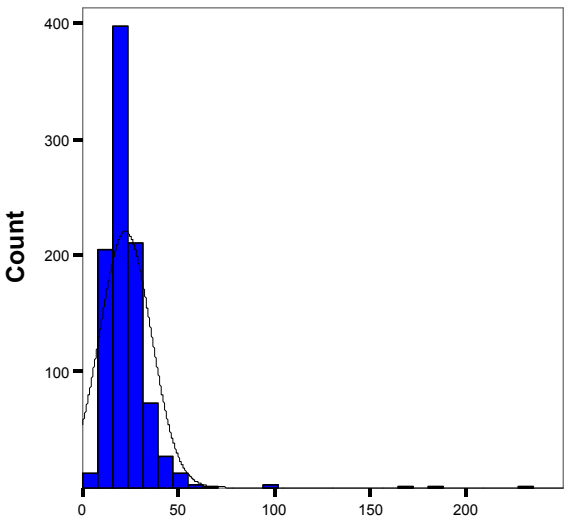
The distributions of plasma Aβ40 and Aβ42 concentrations, after the exclusion of outlier samples, were analysed using the software SPSS v.12 (**Figure 4.3**). The distribution of Aβ40 measurements showed a 22.5-fold variation in values, with a mean of 48 pg/ml (11.1 pM). This increased to a 74-fold range in plasma Aβ42 concentrations, with a mean of 22.5 pg/ml (5 pM). Using the Kolmogorov-Smirnov statistic, both distributions were found to deviate significantly from normality ($P_{A\beta40}$ = 7.6E-25, $P_{A\beta42}$ = 4.8E-75). High kurtosis values also showed that the two distributions were leptokurtic (**Figure 4.3**). This suggested the need to transform the data to a normal distribution, in order to reduce the probability of false-positive results in further analyses. However, conventional normalisation procedures such as natural log, square-root and cube-root transformation methods failed to transform the two distributions.

A



						Kurtosis	
	N	**Minimum**	**Maximum**	**Mean**	**S.D.**	**Mean**	**S.E.**
Plasma Aβ40 Concentration (pg/ml)	958	7.85	176.57	48.03	14.2884	11.7	0.16

B



						Kurtosis	
	N	**Minimum**	**Maximum**	**Mean**	**S.D.**	**Mean**	**S.E.**
Plasma Aβ42 Concentration (pg/ml)	947	3.13	231.71	22.54	13.3817	98.5	0.16

Figure 4.3 Plasma Aβ measurements in general population samples from the island of Vis.

Plasma Aβ concentrations are shown after exclusion of outliers as described in the text. Plasma Aβ40 **(A)** and Aβ42 **(B)** concentrations were measured in EDTA plasma samples using a "sandwich" ELISA developed and optimised by the Younkin laboratory (Suzuki *et al.*, 1994). Tables of descriptive statistics are also presented for the two species of Aβ below each graph. Statistical data was obtained using SPSS v.12.

Therefore, a quantile transformation explained in Section 2.5.6 was applied using Minitab v.12, which successfully transformed the data to normal distributions. Transformed measurements were then used to analyse the effects of possible covariates on plasma Aβ concentrations.

4.2.3 Analysis of plasma Aβ covariates in Vis

The association between possible covariates and concentrations of both plasma Aβ40 and Aβ42 were assessed using a univariate analysis of variance implemented in the software SPSS v.12. Normalised distributions of Aβ40, Aβ42 and Aβ42:Aβ40 ratio were used in the analysis of the covariates sex, age, BMI, smoking and *APOE* alleles.

It was hypothesised that the strong association of *APOE* with LOAD might be explained by the differential affinity of APOE protein variants for binding Aβ in plasma as shown in brain (Strittmatter *et al*., 1993; Mahley & Huang, 1999; Stratman *et al*., 2005). However, binding of APOE to Aβ deposits, in which Aβ is at high concentration, may be quite different to plasma, where Aβ is at very low concentration. No significant associations were detected between any of the *APOE* alleles and plasma Aβ40 (**Table 4.1**), Aβ42 (**Table 4.2**) or Aβ42:Aβ40 ratio (**Table 4.3**). *APOE* locus variation was therefore not a significant covariate for plasma Aβ peptide concentration.

Age was identified as a significant covariate for plasma Aβ40 concentration, with older age resulting in increased plasma Aβ40 (β = 0.028, *P* = 1.3E-28) which explained 12.2% of the variation in the unadjusted trait variance (**Table 4.1**). Age was also found to be a significant covariate for changes in the ratio of plasma Aβ42:Aβ40 concentrations, explaining 4.5% of the total trait variance (**Table 4.3**). Increase in age was found to be associated with a significant decrease in the normalised concentration of Aβ42:Aβ40 in plasma (β = 0.017, *P* = 7.5E-11) (**Table 4.3**).

Smoking was associated with an increase in plasma Aβ40 concentration (β=0.304, *P*=1.5E-04). Smoking was found to explain 1.5% of the total unadjusted Aβ40 variance (**Table 4.1**).

Table 4.1 Aβ40	Mean	χ^2	d.f.	*P*	β (S.E.)	R^2
Sex	-0.70	0.08	1	0.781	-	0.000
Smoking	-0.85	14.40	1	**1.5E-04**	0.304 (0.08)	0.015
Age	-0.70	123.05	1	**1.3E-28**	0.028 (0.00)	0.122
BMI	-0.70	3.57	1	0.058	-	0.002
APOE2	-0.70	0.27	1	0.604	-	0.000
APOE3	-0.70	0.49	1	0.481	-	0.000
APOE4	-0.70	0.17	1	0.675	-	0.000

Mean: -0.701 (±1.24, SD)

Table 4.2 Aβ42	Mean	χ^2	d.f.	*P*	β (S.E.)	R^2
Sex	-0.70	0.07	1	0.793	-	0.000
Smoking	-0.81	6.04	1	**0.014**	0.020 (0.08)	0.006
Age	-0.70	0.81	1	0.368	-	0.000
BMI	-0.70	0.16	1	0.684	-	0.000
APOE2	-0.70	0.01	1	0.917	-	0.000
APOE3	-0.70	0.22	1	0.637	-	0.000
APOE4	-0.70	0.51	1	0.472	-	0.000

Mean: -0.701 (±1.24, SD)

Table 4.3 Aβ42:Aβ40	Mean	χ^2	d.f.	*P*	β (S.E.)	R^2
Sex	-0.70	0.16	1	0.670	-	0.000
Smoking	-0.70	0.01	1	0.933	-	0.000
Age	-0.70	42.38	1	**7.5E-11**	-0.017 (0.003)	0.045
BMI	-0.70	1.79	1	0.181	-	0.000
APOE2	-0.70	0.01	1	0.928	-	0.000
APOE3	-0.70	0.18	1	0.668	-	0.000
APOE4	-0.70	0.43	1	0.513	-	0.000

Mean: -0.701 (±1.24, SD)

Tables 4.1-4.3 Analysis of covariate effects associated with concentrations of Aβ40, Aβ42 and Aβ42:Aβ40 ratio in plasma.

A univariate analysis was used to detect significant associations between changes in plasma concentrations of Aβ40 (**Table 4.1**, 937 individuals), Aβ42 (**Table 4.2**, 928 individuals) and Aβ42:Aβ40 ratio (**Table 4.3**, 890 individuals) and potential covariates sex, age, smoking, BMI and *APOE* alleles. Significant associations are highlighted in bold. d.f. is the degrees of freedom of the χ^2 test. β is the regression coefficient and shows the unit change in the transformed mean QT associated with a unit change in the covariate. R^2 represents the proportion of the total unadjusted phenotypic variance explained. Analysis was carried out using the software package SOLAR (Almasy and Blangero, 1998) after taking the relatedness of individuals into account.

Smoking was also a significant covariate for plasma Aβ42 concentration, explaining 0.6% of the total trait variance (**Table 4.2**). Smoking was found to be associated with a significant increase in the normalised Aβ42 concentration (β = 0.02, P = 0.014) (**Table 4.2**).

4.2.4 Plasma Aβ heritability in Vis

Narrow-sense heritabilities (h^2) were estimated for normalised Aβ40, Aβ42 and Aβ42:Aβ40 concentrations using plasma measurements from 588 individuals in 126 general population pedigrees, collected from the island of Vis. A linear regression model implemented in the software package SOLAR (Almasy & Blangero, 1998) was used for the analysis.

Estimates of narrow-sense heritability (h^2) did not reach significance for any of the traits examined ($h^2_{A\beta40}$ = 0.028 ± 0.114, P = 0.401; $h^2_{A\beta42}$ = 0.168 ± 0.112, P = 0.130; $h^2_{A\beta42:A\beta40}$ = 0.130 ± 0.104, P = 0.101). These findings are discussed in the summary and conclusions section of the chapter.

4.2.5 A genome-wide linkage scan of plasma Aβ in the Vis population

A multipoint genome-wide linkage scan of normalised plasma Aβ40 (N = 952), Aβ42 (N = 942) and Aβ42:Aβ40 (N = 904) was carried out using 810 microsatellite markers typed in DNA samples collected from the villages of Komiza and Vis on the island of Vis. Markers were spaced at approximately 5 cM intervals. Analysis was carried out using the variance components methodology implemented in the software package SOLAR (Almasy & Blangero, 1998). The null hypothesis being tested was that the additive genetic variance due to a putative quantitative trait locus (QTL) equalled zero. The null model therefore did not include a QTL effect, but modelled only the polygenic effects on the phenotype. The null hypothesis of no linkage was tested by comparing the likelihood of the null polygenic model with that of a model that estimated the additive genetic variance at the putative QTL. The significant covariates of age (Aβ40 and Aβ42:Aβ40) and smoking (Aβ40 and Aβ42) were also included in the model.

The results of the linkage scan for all 22 autosomal chromosomes are summarised in **Figures 4.4-4.6**. No multipoint LOD (Z) scores indicative ($Z \geq 3$) or suggestive ($Z \geq 2$) of linkage were observed in this analysis. However, a number of loci reached LOD scores above an arbitrary threshold of 0.6 for the traits examined (**Figures 4.4-4.6)**. These are summarised in **Table 4.4**. For plasma Aβ40, no peaks of $Z \geq 1$ were detected. The peak observed on chromosome 6q23 between markers D6D1656 and D6S281 (LOD = 0.92) was the highest linkage peak detected in the plasma Aβ40 analysis.

Two loci were detected with $Z \geq 1$ on chromosomes 2 and 7 in the analysis of plasma Aβ42. The highest linkage peak was observed on chromosome 2p25 (LOD = 1.19) between markers D2S2211 and D2S391. A linkage peak between markers D7S657 and D7S640 was detected on chromosome 7q21 with a LOD score of 1.09.

Using the ratio of Aβ42:Aβ40 in plasma, no linkage peaks were detected with $Z \geq 1$. All linkage peaks covered large genomic regions with many potential candidate genes and none reached even suggestive significance levels. This therefore emphasised the need for genome-wide association analysis using a high density of single nucleotide polymorphisms. The results of the association analysis of plasma Aβ40, Aβ42 and Aβ42:Aβ40 ratio are presented in the Chapter 5.

4.2.6 Summary and discussion

In this chapter, the results were presented of the biochemical assay and genetic linkage analysis of Aβ in a general population sample from the island of Vis. The two amyloid-β species, Aβ40 and Aβ42, were used as biomarkers of Alzheimer disease. Aβ is the main component of extracellular senile plaques in the brains of patients diagnosed with Alzheimer disease (Glenner & Wong, 1984; Masters *et al*., 1985; Spires & Hyman, 2004). The assembly of Aβ into amyloid fibrils and various toxic intermediates has been implicated as the main pathologic trigger of Alzheimer disease.

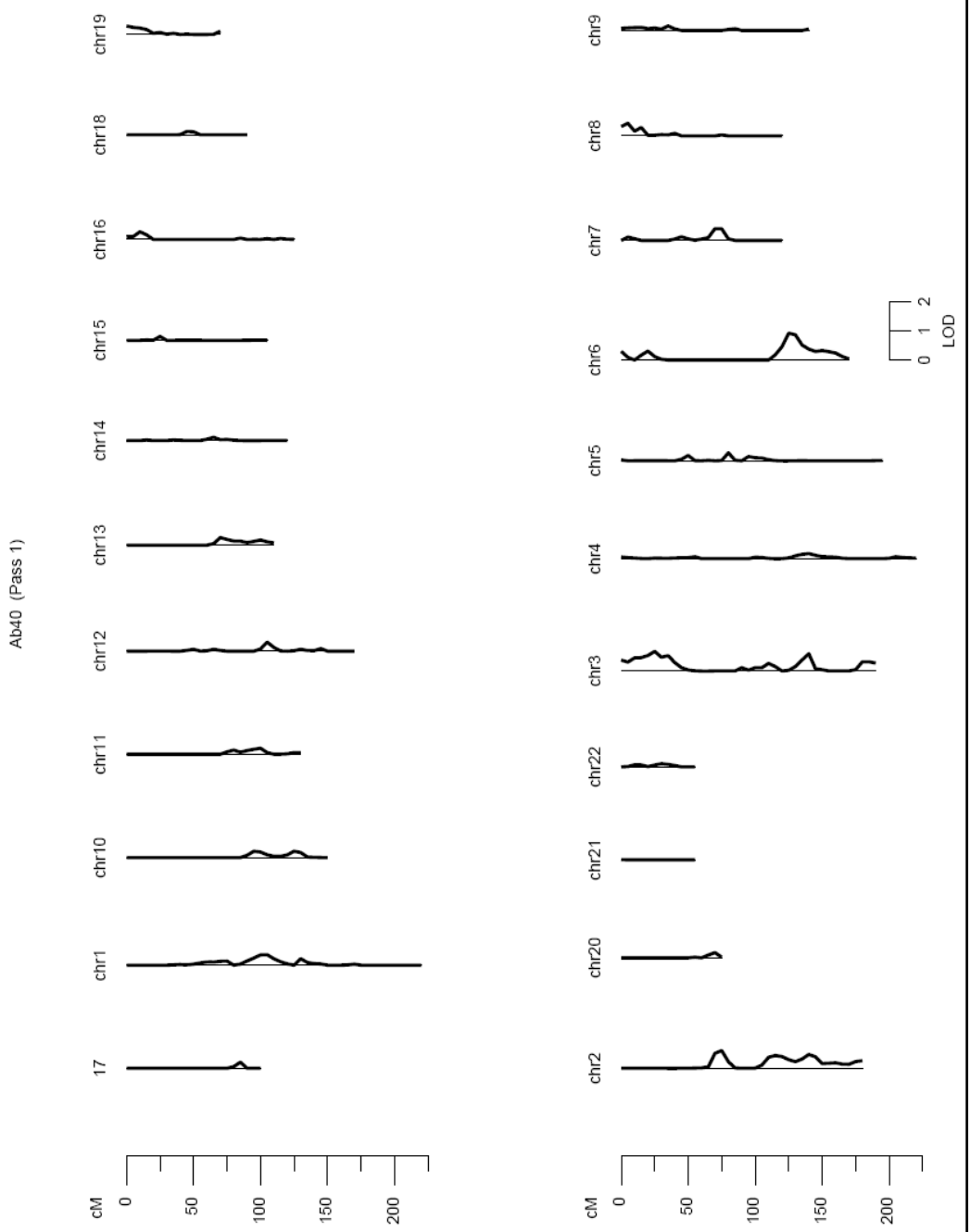


Figure 4.4 Results of plasma Aβ40 genome-wide linkage scan.

LOD scores obtained following a multipoint genome-wide linkage scan with 810 microsatellite markers are presented for each autosomal chromosome. Markers were spaced at approx. 5cM intervals. The analysis used a polygenic model containing smoking and age as significant covariates for plasma Aβ40 levels. The analysis was carried out using the software package SOLAR (Almasy & Blangero, 1998).

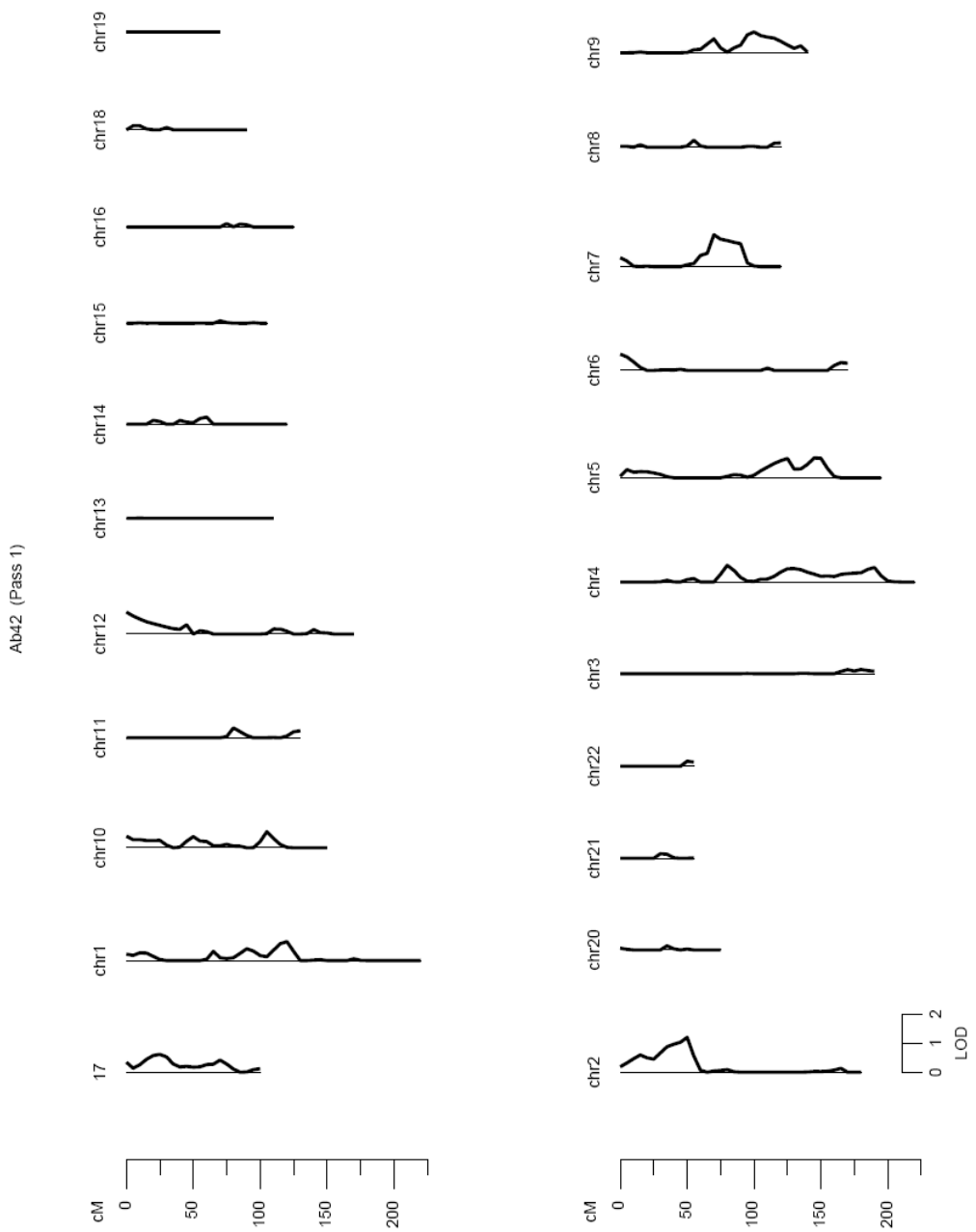


Figure 4.5 Results of plasma Aβ42 genome-wide linkage scan.

LOD scores obtained following a multipoint genome-wide linkage scan with 810 microsatellite markers are presented for each autosomal chromosome. Markers were spaced at approx. 5cM intervals. A polygenic model was used with smoking as a significant covariate for plasma Aβ42 levels. The analysis was carried out using the software package SOLAR (Almasy & Blangero, 1998).

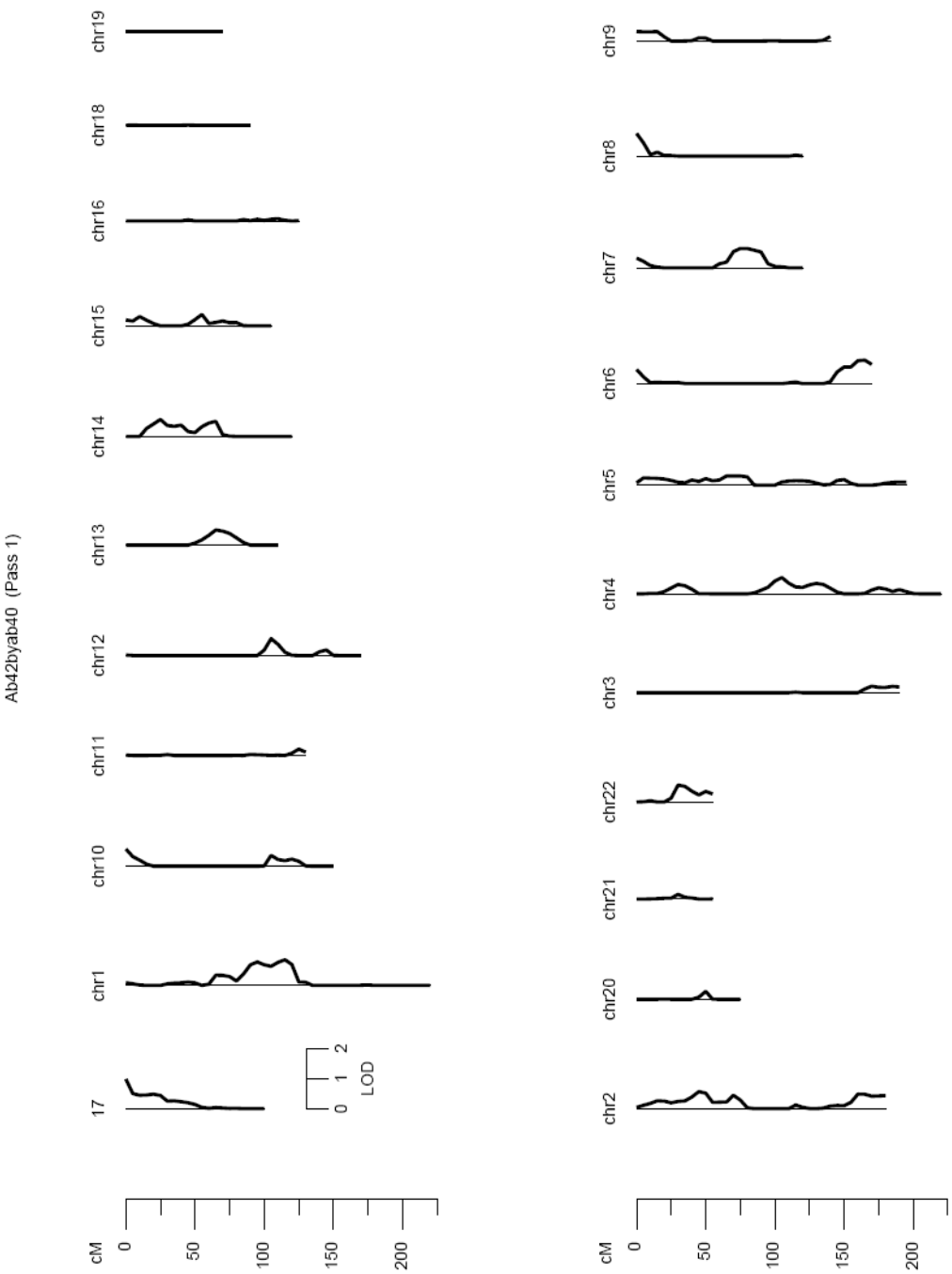


Figure 4.6 Results of genome-wide linkage scan for the ratio of Aβ42:A β40 in plasma.

LOD scores were obtained following a multipoint genome-wide linkage scan with 810 microsatellite markers and are presented for each autosomal chromosome. Markers were spaced at approx. 5cM intervals. A polygenic model was used including age as a significant covariate. The analysis was carried out using the software package SOLAR (Almasy & Blangero, 1998).

Trait	Chromosome	LOD	Marker (from)	Marker (to)	Genetic distance (cM)	Genomic context (bp)
Aβ40	6q23	0.92	D6S1656	D6S281	50.88	132180506-169815699
	3p26	0.67	D3S1270	D3S3567	51.58	1423261-30714032
	2p11	0.60	D2S388	D2S347	6.73	85988497-123966608
Aβ42	2p25	1.19	D2S2211	D2S391	37.82	7421976-46323416
	7q21	1.09	D7S657	D7S640	23.90	92450837-132097253
Aβ42:Aβ40	17p13	0.98	D17S849	D17S927	51.95	379287-32080844
	1p31	0.85	D1S198	D1S412	66.34	66723183-190797799
	6q26	0.78	D6S1599	D6S446	18.90	162810006-170470289
	8p23	0.75	D8S264	D8S550	15.20	2117696-10919323
	7q21	0.64	D7S669	D7S684	35.80	77517087-138001294

Table 4.4 Summary of results from the genome-wide linkage analysis in the Vis population.

Results of the genome-wide multipoint linkage analysis using 810 microsatellite markers in about 900 individuals from the island of Vis. The analysis was carried out for plasma concentrations of Aβ40, Aβ42 and the ratio of Aβ42:Aβ40. Markers were located at approximately 5 cM intervals and the genetic distance between them was estimated using the Haldane mapping function and constructed using the software CRIMAP. The genomic context was based on the NCBI Build 35 assembly.

Although Aβ42 accounts for approximately 10% of secreted Aβ, Aβ42 is the predominant species accumulated in senile plaques in AD brain and appears to be the initially deposited species, hence its plasma concentration may decrease early in the dementia process (Yamazaki *et al*., 1997). Therefore, a high plasma concentration of Aβ40, when combined with a low concentration of Aβ42 (i.e. a low ratio of Aβ42:Aβ40) could indicate an increased risk of dementia as a result of Aβ42 deposition in the brain (van Oijen *et al*., 2006; Graff-Radford *et al*., 2007).

Amyloid-β assay

Plasma concentrations of Aβ40 and Aβ42 were used as surrogate biomarkers for Alzheimer disease. The concentrations of Aβ40 and Aβ42 were measured in 999 plasma samples collected from the island of Vis.

The validity of a commercial ELISA kit (Biosource®) for measurement of Aβ peptides in EDTA plasma was assessed as well as a "sandwich" ELISA developed and optimised for use in human plasma (Suzuki *et al*., 1994) in collaboration with Professor S. G. Younkin (Mayo Clinic, Jacksonville, USA). The Younkin assay proved to be more sensitive for detection of Aβ species in EDTA plasma than the commercial assay and inter and intra-experimental variation were low. The distributions of plasma Aβ40 (mean = 48.03 pg/ml) and Aβ42 (mean = 22.54 pg/ml) showed a 22-fold and 74-fold range of values respectively (**Figure 4.3**). The plasma concentration of Aβ40 (mean 11.1 pM) was unexpectedly low compared with other studies (e.g. 119.2 pM, Ertekin-Taner *et al*., 2008; 44.3 pM, van Oijen *et al*., 2006; 30.7 pM, Mayeux *et al*., 2003) although there is considerable variation between studies. In addition the proportion of total Aβ due to Aβ40 is 69% in the present study, similar to Mayeux *et al*. (2003) (70%) but lower than van Oijen *et al*. (2006) (92%) and Ertekin-Taner *et al*. (2008) (91%). The plasma concentration of Aβ42 (mean 5 pM) is however similar to other studies (4 pM, van Oijen *et al*., 2006; 12 pM, Ertekin-Taner *et al*., 2008; 13 pM Mayeux *et al*., 2003). Such differences could be due to age and other covariates.

Identification of covariates for plasma Aβ

The analysis of covariate effects was carried out after removal of outlier measurements from the Aβ distributions. Plasma Aβ40 and Aβ42 concentrations showed leptokurtic distributions, which were consequently normalised using a quantile transformation.

Age and smoking were identified as significant covariates for plasma Aβ40, explaining 12.2% (P = 1.3E-28) and 1.5% (P = 1.5E-04) of the total trait variance, respectively (**Table 4.1**). Smoking was found to be significantly associated with an increase in plasma Aβ42 concentration, explaining 0.6% of the total trait variance (P = 0.014) (**Table 4.2**).

Published data have shown a reduced density of cortical plaques and elevated nicotinic receptor binding in individuals with an established history of smoking tobacco (Perry *et al*., 1996; Perry *et al*., 2000). Nicotine, the primary alkaloid in tobacco, has been shown to partially prevent the neurotoxicity of Aβ (Zamani & Allen, 2001) possibly by disruption of Aβ fibrils leading to reduced peptide aggregation (Ono *et al*., 2002; Dickerson & Janda, 2003). Furthermore, levels of soluble and insoluble Aβ40 and Aβ42 in frontal cortex and Aβ40 in temporal cortex and hippocampus have been shown to be significantly decreased in smoking compared to non-smoking individuals/patients with Alzheimer disease (Hellström-Lindahl *et al*., 2004). The association of cigarette smoking with increased levels of Aβ40 and Aβ42 in plasma may therefore, in agreement with published data, indicate release of Aβ species from the brain as a result of the disruption of amyloid fibrils.

Finally, an increase in age was associated with a decrease in the ratio of Aβ42:Aβ40 in plasma (P = 7.5E-11), explaining 4.5% of the total trait variance (**Table 4.3**). Age has long been known to be a major confounder in susceptibility to AD and is associated with an increased incidence and quantity of neurofibrillary tangles and senile plaques (Morimatsu *et al*., 1975; Younkin, 1998). The concentrations of Aβ40 and Aβ42 in plasma have also been shown to increase significantly with age until early in the disease process when Aβ42 starts accumulating in the brain (Fukumoto *et al*., 2003; Mayeux *et al*., 2003). A significant association between plasma Aβ40 concentration and age has been

reported in a population of non-demented individuals (Mayeux *et al*., 1999; Mehta *et al*., 2000). An increase in the mean plasma Aβ40 in the general population of Vis is therefore consistent with the findings of Mayeux *et al*. and Mehta *et al*. However, it has also been suggested that elevated plasma concentration of Aβ42 can be detected several years before the onset of symptoms of late-onset AD (Mayeux *et al*., 1999). The lack of association between plasma Aβ42 and age in the Vis population is in contrast to this finding. On the other hand, the association between older age and a decrease in the plasma Aβ42:Aβ40 ratio could be indicative of a reduction in plasma Aβ42 concentration. This is unlikely however, since no direct association between plasma Aβ42 and age was observed in this population.

No significant associations were observed between concentrations of plasma Aβ40, Aβ42 and Aβ42:Aβ40 and either sex or BMI in the Vis data set. However, elevated plasma Aβ42 concentrations have been reported in females with mild cognitive impairment (Assini *et al.*, 2004).

No significant associations were observed between the three *APOE* alleles and concentrations of Aβ species in plasma (**Tables 4.1-4.3**). However, the APOE*E4 isoform has been shown to form a stable complex with Aβ peptide *in vitro* (Strittmatter *et al*., 1993). In contrast, lipidated APOE*E3 has been found to bind Aβ peptide with a 20-fold higher affinity than lipidated E4 (Stratman *et al*., 2005). Also, studies have confirmed the presence of the APOE*E4 protein in brain lesions and the role of the human APOE in amyloid deposition in the AD mouse model by injecting APOE isoforms into the mouse plasma (Bales *et al*., 1999). The hypothesis that APOE protein variants regulate Aβ species in plasma and hence affect susceptibility to LOAD was therefore rejected. However, Ertekin-Taner *et al*. (2008) found significantly reduced Aβ42 in *APOE*E4* homozygotes compared with other genotypes, which was not evident for Aβ40. However, the low allele frequency of the *E4* allele in the present study will have reduced the power to detect this effect of *APOE* (*E4* homozygotes represent 0.5-0.9% of the Vis population).

Heritability of plasma Aβ

Narrow-sense heritability estimates for plasma Aβ40, Aβ42 and Aβ42:Aβ40 concentrations did not reach significance in the Vis population samples. This does not necessarily reflect the lack of a genetic component. Low estimates of h^2 could be due to lack of power as a result of small sample size or the presence of many distant relationships. A low narrow-sense heritability could also be consistent with the presence of recessive alleles which have a strong effect in rare homozygotes on plasma Aβ concentrations. Recessive loci of large effect would result in non-significant estimates of narrow-sense heritability, which only accounts for additive genetic variance. On the other hand, estimation of broad-sense heritability, based on the total genetic variance (additive, dominance and epistatic variance) would not be possible in the Vis sample set due to the small sibship size in the pedigrees.

Genome-wide linkage analysis of plasma Aβ

The genome-wide multipoint linkage analysis did not identify any significant linkage peaks using 810 microsatellite markers in over 900 individuals from the general population of Vis island. Low power for linkage mapping of QTL contributing less than 3% of the variance, which comprise the majority of the QTL identified to date in complex traits, was found using the genetic power calculator (Purcell *et al*., 2003), partly due to the small sample size of most pedigrees. None of the markers showed suggestive evidence of linkage with LOD score values ≥ 2. The highest LOD score for plasma Aβ40 concentration was detected on 6q23 (Z = 0.92). The highest LOD score in the Aβ42 linkage analysis was found on chromosome 2p25 (Z = 1.19) followed by 7q21 (Z = 1.09). The highest LOD score for plasma Aβ42:Aβ40 ratio was observed on 17p13 (Z = 0.98). Results of the linkage analysis have been discussed in the Discussion chapter. In short, failure of linkage can be either due to the small effect size or low frequencies of the underlying additive genetic factors influencing plasma Aβ concentrations, consistent with the non-significant Aβ heritabilities (Sham *et al*., 2000; Dawn & Barrett, 2005). In addition, the small size of the pedigrees reduces the power of linkage analysis to detect typical QTL effects (Blangero, 2004).

In conclusion, the analyses of plasma Aβ covariates, plasma Aβ heritabilities and a genome-wide linkage scan of plasma Aβ were carried out in a general population sample. The results showed that age and smoking were found to be significantly associated with concentrations of Aβ40 and Aβ42 in plasma. The possibility of an interaction between APOE protein variants and Aβ species in plasma was excluded in a univariate analysis of association between *APOE* alleles and plasma Aβ concentration. Since multipoint linkage analysis of plasma Aβ did not identify either suggestive or significant linkage peaks due to the low power of linkage to detect typical QTL effects, a high-resolution genome-wide association analysis was carried out to identify genomic loci of smaller effect associated with plasma Aβ.

CHAPTER 5

GENOME-WIDE ASSOCIATION ANALYSIS OF PLASMA Aβ

5.1 Introduction

In this chapter, the results of a whole-genome association analysis of plasma amyloid-β (Aβ), using the Illumina Human Hap300 Array are described. Concentrations of Aβ40, Aβ42 and the ratio of Aβ42:Aβ40 in 966 plasma samples collected from the Croatian island of Vis were used as intermediate phenotypes in a quantitative trait locus (QTL) association study. Plasma Aβ measurements were carried out using a "sandwich" ELISA as described in the previous chapter.

Mutations in three genes, *APP*, *PSEN1* and *PSEN2* account for about 70% (Campion *et al*., 1999) of patients with autosomal dominant familial Alzheimer disease (FAD) (St George-Hyslop *et al*., 1987). Although FAD accounts for less than 1% of all AD cases worldwide, the identification of rare mutations in these genes led to the pathophysiological hypothesis of the amyloid cascade (**Figure 1.7**). Genetic variation is believed to play a significant role in the Alzheimer disease process, since heritability of AD has been estimated to be around 60% (Gatz *et al.*, 1997). In addition, *APOE*E4* has been identified as a common genetic susceptibility variant influencing late-onset Alzheimer disease (LOAD). It has been suggested that there are at least four other genes of large effect similar to *APOE*E4* and many other genes of smaller effect in Alzheimer disease (Daw *et al*., 2000).

Linkage to chromosome 10q22 has previously been detected in extended LOAD pedigrees, containing affected individuals with high plasma Aβ42 and/or Aβ40 concentrations (Ertekin-Taner *et al*., 2000). The *IDE* gene, encoding insulin degrading enzyme which is known to degrade Aβ, has been shown to be associated with increased risk of LOAD (odds ratio of 2.7), possibly by influencing plasma Aβ42 concentration ($P = 0.02$) (Prince *et al*., 2003; Ertekin-Taner *et al*., 2004). The *PLAU* gene, which encodes a urokinase-type plasminogen activator, also resides in the 10q22 region. Polymorphisms in *PLAU* have been shown to influence susceptibility to LOAD in a study involving three independent series, each of which showed significant association between *PLAU* and plasma Aβ42 with an odds ratio of 1.2 ($P = 0.0006$) (Ertekin-Taner *et al*., 2005).

In this thesis, direct associations between plasma Aβ concentration and genomic polymorphisms were assessed using a QT genome-wide association

approach. Circulating Aβ concentrations were used as surrogate biomarkers for AD with the aim of identifying genomic loci which influence the concentrations of these peptides in plasma, which could influence an individual's susceptibility to LOAD. An understanding of the genetic influences on intermediate traits such as Aβ peptides may help to identify new pathways involved in Alzheimer disease process.

5.2 Results

5.2.1 Genome-wide association analysis of plasma Aβ in Vis

Genome-wide association analysis of QTL affecting concentrations of Aβ40, Aβ42 and Aβ42:Aβ40 in plasma was carried out in 966 individuals from the island of Vis. The Illumina Human Hap300, based on the HapMap data Release 21A and consisting of 317,503 single nucleotide polymorphisms (SNP), was used in the association analyses. The genome-wide association tool set PLINK version 0.99s (Purcell *et al*., 2007) was used to perform the pedigree-based QT association analysis (Aulchenko *et al*., 2007).

Outlier values within each of the Aβ40 and Aβ42 measurements were excluded as described before. The distributions of Aβ40, Aβ42 and the ratio of Aβ42:Aβ40 were then normalised by a ranking transformation method, as described in Section 2.5.6. The residuals were first obtained from a regression model of Aβ40, Aβ42 and Aβ42:Aβ40 concentrations after adjusting for the relatedness of individuals, covariates and fixed effects. These residuals were then used as the dependent variable in a linear regression analysis for estimating the effect of each SNP on plasma Aβ peptides. Finally, SNPs showing evidence of association to plasma Aβ were analysed using a full Measured Gentoype (MG) model to obtain fixed effect sizes and *P*-values (Section 2.5.7.4) (Aulchenko *et al*., 2007).

5.2.1.1 QTL association analysis of plasma Aβ40

The concentrations of Aβ40 measured in plasma from 904 individuals were used to identify associated QTL in a genome-wide scan. The results of the genome-wide association analysis of Aβ40 using 317,503 SNPs are presented in **Figure 5.1**.

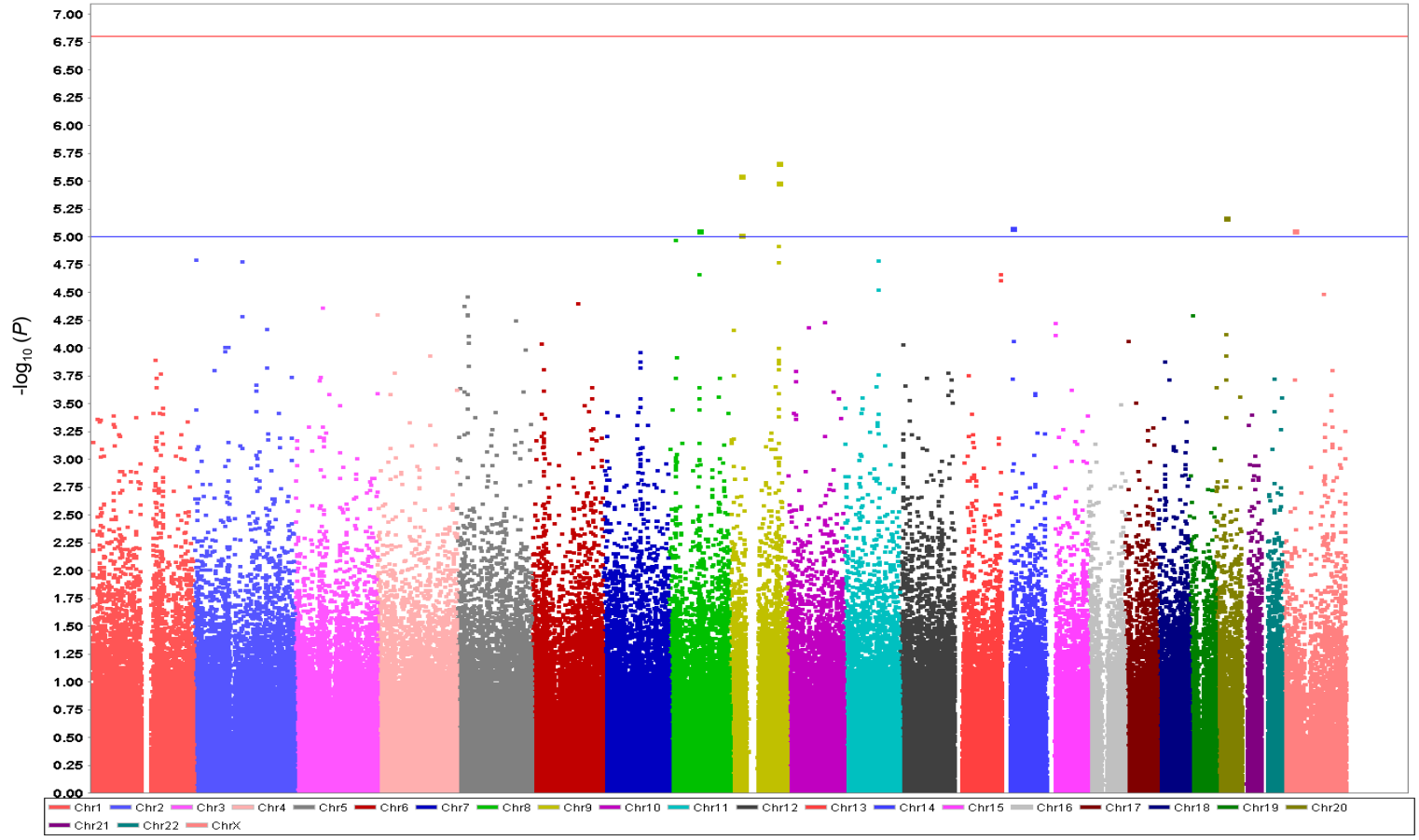


Figure 5.1 Results of the pedigree-based genome-wide quantitative trait locus association analysis of plasma Aβ40 in Vis.

Results of the QT association test of 308,140 markers are shown for each chromosome. Analysis was carried out using normalised plasma Aβ40 measurements in 966 individuals from the island of Vis, after correcting for age, smoking and relatedness. Unadjusted $\log_{10}$ transformed *P*-values for association are presented on the Y-axis. A threshold for genome-wide significance (red line) was using a Bonferroni correction. Suggestive significance is marked by the blue line. The figure was obtained from Haploview version 4.0 (Barrett *et al*., 2005).

A total of 8,984 SNPs were removed due to low genotyping rate (< 90%) and 379 SNPs were removed due to low minor allele frequency (< 1%) in the population, leaving 308,140 in the final analysis.

None of the markers used in the analysis passed the threshold of significance after a Bonferroni correction for multiple testing ($P \leq$ 1.0E-07). However, nine SNPs reached suggestive significance ($P_{association} \leq$ 1.6E-05) and one other showed potential association. The criteria for a potential association were based on the presence of at least one SNP with $P \leq$ 1.0E-04 and/or at least three SNPs with $P \leq$ 1.0E-03 within a 200 kb region.

Based on the above criteria, suggestive and potential associations with plasma Aβ40 were observed on chromosomes 8p, 8q, 9p, 9q, 14q, 20p and Xp. A summary of the SNPs at these loci is shown in **Table 5.1**. Two of the most significant associations with plasma Aβ40 were observed in chromosomal region 9q33 with the Astrotactin 2 (*ASTN2*) SNPs rs2900131 (P = 2.1E-06) and rs7866525 (P = 3.2E-06), which explained 2.4% and 2.3% of the total trait variance respectively using the full MG model. Both SNPs are relatively common in this population ($MAF_{rs2900131}$ = 0.495, $MAF_{rs7866525}$ = 0.493) and lie within intron 16 of the 989.8 kb *ASTN2* gene. The effect sizes associated with the minor allele at both SNPs, represented by the β values in **Table 5.1** were protective, resulting in a significant reduction in plasma Aβ40 concentration ($\beta_{rs2900131}$ = -0.23 ± 0.05 (SE), $\beta_{rs7866525}$ = -0.22 ± 0.05 (SE)). Analysis of the association peak at 9q33 revealed a cluster of associated SNPs which spanned a region of about 500 kb within the 3′ end of the *ASTN2* gene showing high levels of LD (**Figure 5.2**). Both rs2900131 and rs7866525 were found to be in strong linkage disequilibrium (r^2 > 0.80) with rs2295741 (SNP 28, **Figure 5.3**), which was 50 bp upstream of *ASTN2* exon 15 close to the exon-intron junction.

ASTN2 is a transmembrane protein with several domains, including three epidermal growth factor (EGF)-like domains, a membrane attack complex/perforin (MACPE) domain and one fibronectin type 3 (FN3) domain, which is mainly encoded by the 5′ end of the gene. Four isoforms of ASTN2 have been identified, namely A, B, C and D, which differ in their transcription start site.

SNP	Minor allele	Chromosome	Position (bp)	MAF	N_{chrom}	$P_{association}$	R^2	P_{HWD}	β	S.E.	Genomic context
rs2900131	G	9q33	116,426,248	0.495	1932	2.1E-06	0.024	0.604	-0.23	0.05	*ASTN2* intron 16
rs10968018	G	9p21	27,617,785	0.083	1918	2.8E-06	0.024	0.671	-0.42	0.09	81 kb upstream of *C9ORF72*
rs7866525	G	9q33	116,437,518	0.493	1932	3.2E-06	0.023	0.331	-0.22	0.05	*ASTN2* intron 16
rs761865	A	20p11	24,021,113	0.419	1928	6.6E-06	0.022	0.894	0.22	0.05	Intergenic region
rs1461554	G	14q12	24,756,156	0.250	1932	8.1E-06	0.022	0.863	-0.25	0.06	Intergenic region
rs1473522	G	8q13	71,996,994	0.205	1864	8.6E-06	0.022	0.316	-0.28	0.06	Intergenic region
rs12009327	G	Xp21	31,251,637	0.137	1473	8.6E-06	0.022	0.474	-0.27	0.06	*DMD* intron 39
rs3913107	G	9p21	27,613,558	0.071	1926	9.4E-06	0.021	0.617	-0.42	0.10	77 kb upstream of *C9ORF72*
rs7829617	T	8p22	16,442,565	0.391	1846	1.0E-05	0.022	0.032	0.21	0.05	Between *MSR1* and *FGF20*
rs10504478	C	8q13	71,707,472	0.199	1898	2.1E-05	0.020	0.221	-0.26	0.06	3 kb downstream of *LACTB2*

Table 5.1 Summary of the results of a genome-wide association analysis of plasma Aβ40 in the Vis population.

SNPs showing suggestive ($P \leq 1.0E-05$) or potential ($P \leq 1.0E-04$) associations in the genome-wide association analysis of plasma Aβ40 in 966 individuals from Vis. Analysis was based on an additive model of the minor allele after correcting for age, smoking and relatedness. Minor allele frequencies (MAF) and the number of chromosomes (*N*) in the population sample are also shown. R^2 represents the proportion of the phenotypic variance explained by each SNP. The direction of the effect of the minor allele is represented by β, the regression coefficient of the trait and its standard error (S.E.). Values of β are arbitrary as these are based on the transformed distributions of plasma Aβ40 levels. All markers were found to be in Hardy-Weinberg equilibrium. Analysis was carried out using the software package PLINK (Purcell *et al*., 2007).

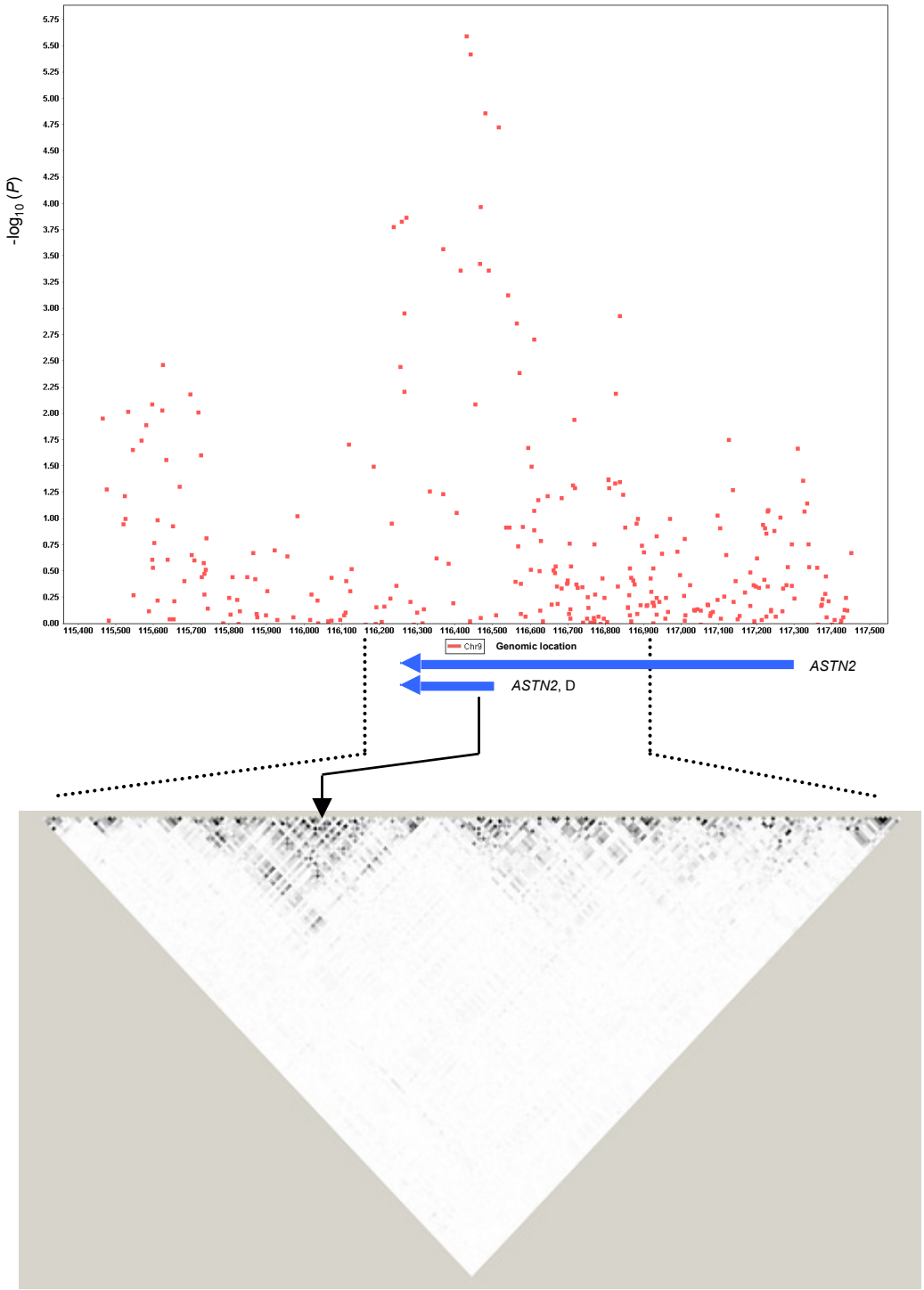


Figure 5.2 Analysis of the chromosomal region 9q33 showing suggestive evidence of association with plasma Aβ40.

The results of the genome-wide association with plasma Aβ40 concentration in the chromosomal region 9q33. The 2 Mb region spanning the SNP rs2900131 showing suggestive association with plasma Aβ40 is shown above. Analysis of linkage disequilibrium (r^2) (below) shows several LD blocks in the 0.9 Mb region containing the cluster of suggestively associated SNPs, with decreasing shades of black representing reductions in the level of pairwise LD. The black arrow shows the approximate position of the SNP rs2900131. The positions of the full-length *ASTN2* and the *ASTN2* isoform D-coding genes are also shown to scale (blue arrows show the direction of transcription). Analysis was carried out in 492 unrelated individuals from Vis using the software Haploview.

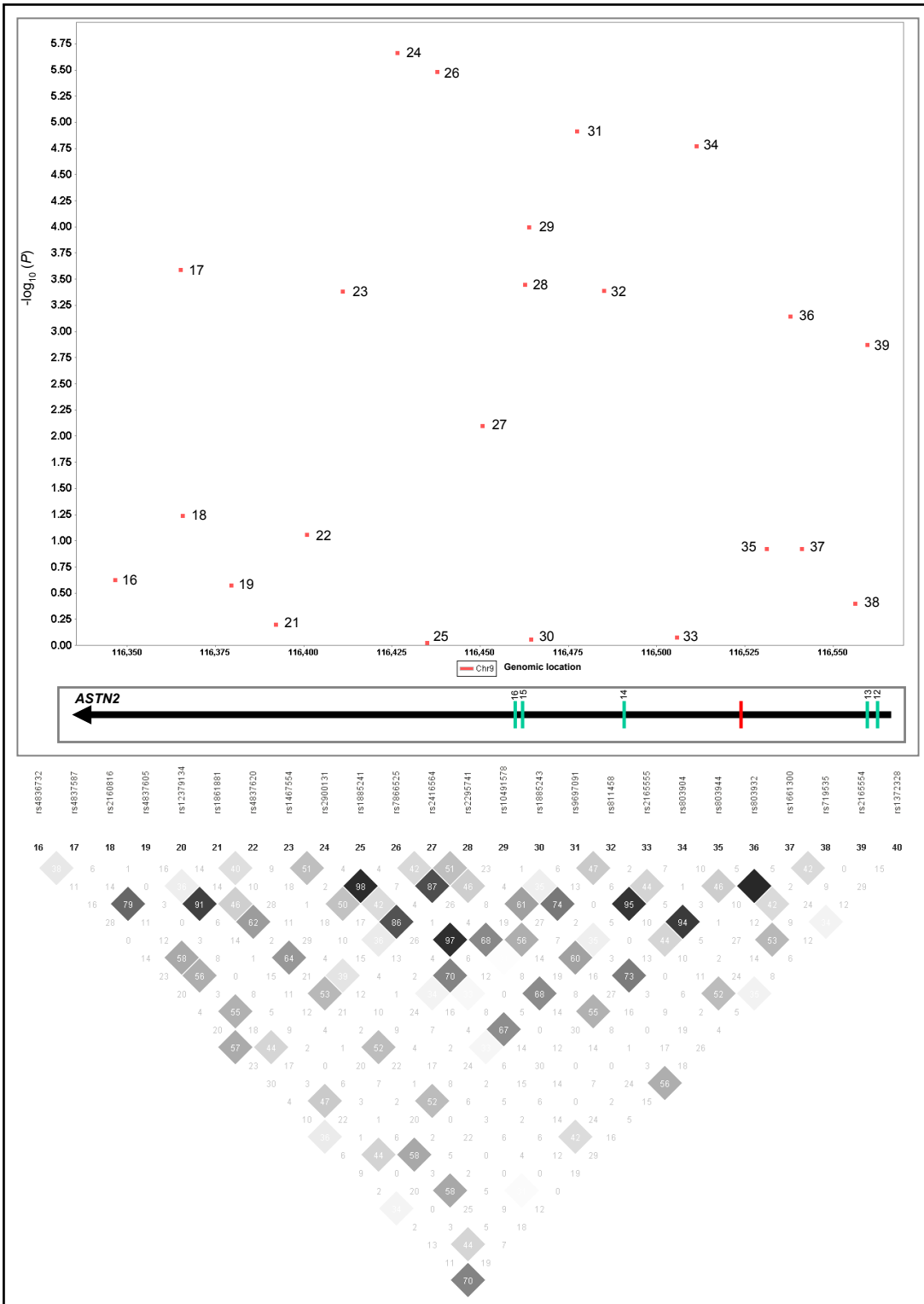


Figure 5.3 Analysis of plasma Aβ40 association and linkage disequilibrium between SNPs at the 9q33 locus.

The 9q33 locus spanning part of the *ASTN2* gene (drawn below the association figure, to scale), shows suggestive association with plasma Aβ40 in the Vis population. Levels of pairwise linkage disequilibrium (LD) shown in terms of the squared correlation (r^2) are also presented. Decreasing shades of black represent reductions in the level of pairwise LD. Analysis of LD was carried out in 492 unrelated individuals from Vis using Haploview. Vertical green bars mark exons 12-16 of the full length *ASTN2*. The red bar marks the start position for the *ASTN2* isoform D. Direction of transcription is shown by the black arrow head.

The region of high LD containing the associated SNPs was found near the transcription start site of the *ASTN2* isoform D (**Figure 5.2**). Furthermore, the nearest SNP to the *ASTN2* isoform D start sites (vertical red line, **Figure 5.3**), rs803904, was found to be in strong LD ($r^2 > 0.60$) with the two most strongly associated SNPs, rs2900131 and rs7866525 (**Figure 5.3**).

Two markers on chromosome 9p21 (rs10968018 and rs3913107) also reached the suggestive significance threshold ($P_{rs10968018}$ = 2.8E-06, $P_{rs3913107}$ = 9.4E-06). Each SNP accounted for about 2.2% of the phenotypic variance and both were associated with a significant reduction in plasma Aβ40 concentrations, as shown by the value of β (**Table 5.1**). The two SNPs lie within 4 kb of each other and about 80 kb away from a hypothetical gene, *C9orf72*.

An intronic SNP, rs12009327, within intron 39 of the Dystrophin (*DMD*) gene in chromosomal region Xp21 showed association with plasma Aβ40, reaching suggestive significance (*P* = 8.6E-06). The nearest Illumina SNP to rs12009327 is rs120011023, which is located 10.9 kb downstream and showed weaker association ($P_{association}$ = 1.9E-04). *DMD* rs12009327 is located about 4 kb downstream of exon 39 in a region of "high regulatory potential" based on the scores computed from alignments of DNA sequences from human, chimpanzee, macaque, mouse, rat, dog, and cow (UCSC Human Genome Browser, March 2006 assembly). Regulatory potential scores are obtained by comparing the frequencies of short alignment patterns between known regulatory elements and neutral DNA (King *et al*., 2005; Kolbe *et al*., 2004).

Other markers showing suggestive association with plasma Aβ40 concentration were found to be situated in intergenic regions (rs761865, rs1461554 and rs1473522), often at a considerable distance from the nearest genes (rs7829617 and rs10504478) as shown in **Table 5.1**. Two of these SNPs, rs10504478 (showing potential association, *P* = 2.1E-05) and rs1473522 (showing suggestive association, *P* = 8.6E-06) were situated at the same locus in chromosomal region 8q13. These two markers are about 300 kb apart and SNP rs10504478 was found to be located approximately 3 kb downstream of the Lactamase Beta 2 (*LACTB2*) gene.

5.2.1.2 QTL association analysis of plasma Aβ42

Plasma Aβ42 concentrations were measured in 999 individuals, as described previously. Aβ42 has been reported to be the major constituent of neural plaques in the Alzheimer disease brain. Normalised plasma Aβ42 measurements were used to carry out a genome-wide QTL association analysis which aimed to detect associated genomic loci.

Genome-wide association analysis using 308,140 SNPs revealed three loci (4 SNPs) which passed the threshold for suggestive significance ($P_{association} \leq$ 1.0E-05). These loci were located in chromosomal regions 2q32, 5q14 and Xp11 (**Figure 5.4**). Three more loci (5 SNPs) were detected which showed potential association with plasma Aβ42 based on a SNP cluster with $P \leq$ 1.0E-04. These associations were observed in chromosomal regions 9q33, 15q21 and 20q13.

The most significant association was observed at rs2386421 (P = 5.7E-07) located in intron 3 of the EGF-like repeats and Discoidin I-like domains 3 (*EDIL3*) gene in chromosomal region 5q14 (**Table 5.2**). This polymorphism accounted for 2.8% of the phenotypic variance and the minor allele was found to be associated with a significant increase in Aβ42 concentration. The 400 kb *EDIL3* gene encodes an integrin-ligand protein with an EGF-like and discoidin I-like domain thought to be involved in vascular development. However, rs2386421 was the only SNP detected at this locus, hence it requires replication.

The association peak in chromosomal region 2q32 consisted of two SNPs which reached the threshold for suggestive significance (rs1113932, P = 5.5E-06 and rs4267464, P = 8.7E-06). The two SNPs are located in highly conserved regions predicted to have a strong "regulatory potential" (King *et al*., 2005; Kolbe *et al*., 2004; UCSC Human Genome Browser, March 2006 assembly) in intron 2 of the Transmembrane protein with EGF-like and two Follistatin-like domains 2 (*TMEFF2*) gene.

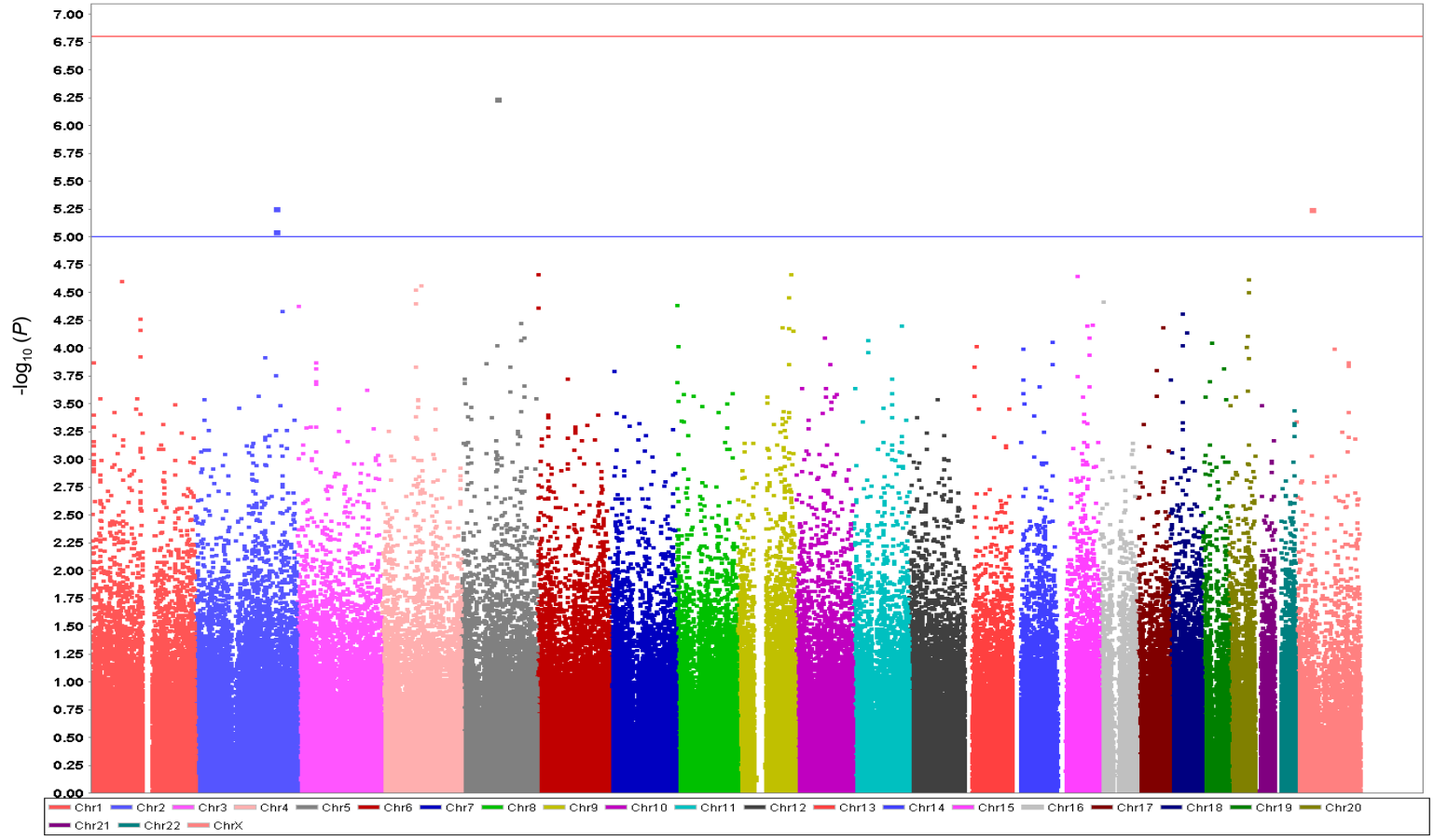


Figure 5.4 Results of the genome-wide quantitative trait locus association analysis of plasma Aβ42 in the Vis population.

Results of the genome-wide QTL association test of 308,140 SNPs with plasma Aβ42 concentration are shown for each chromosome. The analysis was carried out using normalised plasma Aβ42 measurements in 966 individuals from the island of Vis after correcting for smoking and relatedness. Unadjusted $\log_{10}$ transformed *P*-values for association are presented on the y-axis. The threshold for genome-wide significance (red line) was set using a Bonferroni correction for multiple testing. Suggestive significance is marked by the blue line. Figure obtained from Haploview.

SNP	Minor allele	Chromosome	Position (bp)	MAF	N_{chrom}	$P_{association}$	R^2	P_{HWD}	β	S.E.	Genomic context
rs2386421	C	5q14	83556235	0.112	1858	5.7E-07	0.028	0.069	0.42	0.08	*EDIL3* intron 3
rs1113932	C	2q32	192877859	0.118	1924	5.5E-06	0.023	0.214	0.36	0.08	*TMEFF2* intron 2
rs5917513	A	Xp11	37564759	0.450	1524	5.6E-06	0.023	0.549	-0.19	0.04	Gene desert
rs4267464	T	2q32	192875548	0.118	1922	8.7E-06	0.022	0.214	0.36	0.08	*TMEFF2* intron 2
rs2067885	G	15q21	48009237	0.303	1930	2.2E-05	0.020	0.760	0.23	0.05	*ATP8B4* intron 16
rs6122718	T	20q13	46779632	0.278	1918	2.4E-05	0.020	0.572	0.24	0.06	*PREX1* intron 4
rs6095239	A	20q13	46730946	0.309	1932	3.1E-05	0.019	0.545	0.23	0.05	*PREX1* intron 11
rs10513414	C	9q33	122685262	0.068	1932	3.4E-05	0.019	0.614	-0.42	0.10	6 kb downstream of *RC3H2*
rs3739836	T	9q33	122763069	0.063	1928	6.5E-05	0.018	1.000	-0.42	0.10	*ZBTB26* intron 1

Table 5.2 Summary of the results of whole-genome association analysis of plasma Aβ42 in the Vis population.

Details of the associated SNPs observed in a genome-wide association analysis of plasma Aβ42 in 966 individuals from Vis. The analysis was carried out based on an additive model of the minor allele after correcting for smoking and relatedness. SNPs showing suggestive significance ($P_{association} \leq$ 1.0E-05) are shown above the dotted line and those showing "potential" association (see text) are shown below the dotted line. Minor allele frequencies (MAF) and the number of chromosomes (*N*) used from the population sample are also shown. R^2 represents the proportion of the phenotypic variance explained by each SNP. The direction of effect exerted by the minor allele of each SNP is represented by β, the regression coefficient and its standard error (S.E.). Values of β are arbitrary as these are based on the transformed distributions of plasma Aβ42 levels. All markers were found to be consistent within Hardy-Weinberg equilibrium. Analysis was carried out using the software package PLINK (Purcell *et al*., 2007).

Both associated SNPs in *TMEFF2* explain 2% of the underlying phenotypic variance and the minor alleles were associated with an increase in Aβ42 concentration in plasma ($\beta_{rs1113932}$ = 0.36 ± 0.08 SE; $\beta_{rs4267464}$ = 0.36 ± 0.08 SE) (**Table 5.2**). Analysis of LD at the 2q32 locus containing rs1113932, showed that this and two other SNPs, rs4267464 and rs2356960, were located in a region of high LD at the 5′ end of *TMEFF2* (**Figure 5.5**, SNPs 14-18).

A SNP on chromosome Xp11, rs5917513 (*P* = 5.6E-06) also reached suggestive genome-wide significance and the minor allele was found to be associated with a significant increase in plasma Aβ42 concentration (β = 0.19 ± 0.04 SE) (**Table 5.2**). Rs5917513 is an intergenic SNP situated in a non-conserved region 150 kb upstream of the Dynein, Light chain, Tctex-type 3 (*DYNLT3*) gene, encoding a T-complex associated testis-expressed protein, and the Synaptotagmin-Like 5 (*SYTL5*), gene encoding a synaptotagmin-like protein.

A peak of association consisting of rs2067885 (*P* = 2.2E-05) and rs12916154 (*P* = 1.7E-04) showed potential association at a locus within chromosomal band 15q21 (**Table 5.2**). The most strongly associated SNP is located in intron 16 of the ATPase Class I type 8B member 4 (*ATP8B4*) gene in a region of high evolutionary conservation. The latter is based on the Multiz alignment of DNA sequences from 17 vertebrates, including mammalian, amphibian, bird, and fish species using a phylogenetic hidden Markov model (Siepel *et al.*, 2005; UCSC Human Genome Browser, March 2006 assembly). This SNP explains 2.0% of the phenotypic variance and the minor allele results in a significant increase in Aβ42 plasma concentration (β = 0.23 ± 0.05 SE) (**Table 5.2**). The 15q21 locus has previously been reported to be associated with risk of LOAD, in a case/control association study (Li *et al*., 2008). The LOAD study identified an intergenic SNP (rs10519262) between *ATP8B4* and *SLC27A2*, which is in strong LD with *ATP8B4* SNP rs7176805 in a distal promoter region. However, SNP rs10519262 was not found to be in strong LD with either of the above polymorphisms in our study or in the HapMap database.

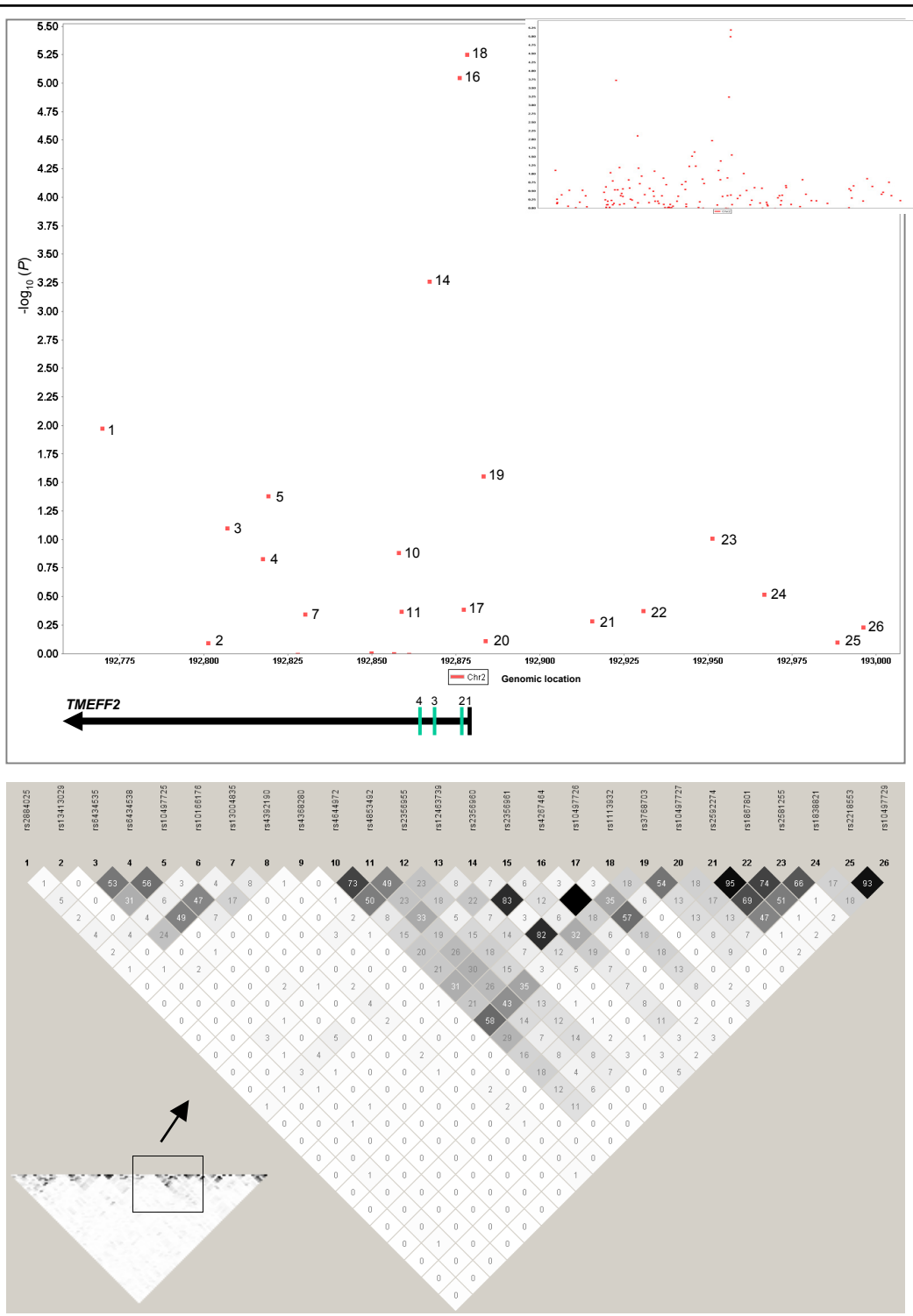


Figure 5.5 Analysis of plasma Aβ42 association and linkage disequilibrium between SNPs at the 2q32 locus.

Association of Aβ42 plasma at the 2q32 locus spanning the 5′ end of the *TMEFF2* gene (drawn below the association figure, to scale). Levels of pairwise linkage disequilibrium (LD) shown in terms of the squared correlation (r^2) are also presented, showing a region of strong LD at the association peak. Decreasing shades of black represent reductions in the level of pairwise LD. Analysis of LD was carried out in 492 unrelated individuals from Vis using Haploview. Vertical bars mark exons 1-4 of *TMEFF2*. The direction of transcription is shown by the black arrow head.

A peak on chromosome 20q13 also showed potential association with Aβ42 plasma concentration (**Figure 5.4**). Two markers, rs6122718 (*P* = 2.4E-05) and rs6095239 (*P* = 3.1E-05), both located in the Phosphatidylinositol 3,4,5-Trisphosphate-dependent RAC Exchanger 1 (*PREX1*) gene were found to account for 2.0% of the phenotypic variance in the Vis population (**Table 5.2**). The protein encoded by this gene acts as a guanine nucleotide exchange factor for the RHO family of small GTP-binding proteins.

A group of SNPs at a 9q33 locus showed potential association with Aβ42 concentration in plasma (**Figure 5.4**). Markers rs10513414 (*P* = 3.4E-05) and rs3739836 (*P* = 6.5E-05) were the most significantly associated SNPs, accounting for at least 1.8% of the phenotypic variance (**Table 5.2**). SNP rs10513414 is located about 6 kb downstream of the Ring Finger and CCCH-type Zinc Finger Domains 2 (*RC3H2* or *MNAB*) gene, encoding a membrane-associated DNA binding protein. SNP rs3739836 is situated in intron 1 of the Zinc Finger and BTB Domain containing 26 (*ZBTB26*) gene, encoding a zinc-finger protein. The associated SNPs cover a region of about 425 kb, which contains many other genes, including another zinc-finger protein coding gene (*ZNF482*), two olfactory receptors (*OR5C1* and *OR1K1*), a G Protein-coupled Receptor (*GPR21*), a RAB GTPase Activating Protein (*RABGAP1*) and a Phosphoducin-Like gene (*PDCL*) gene.

5.2.1.3 QTL association analysis of the ratio of Aβ42:Aβ40 in plasma

The plasma concentrations of Aβ42 and Aβ40 were measured in 999 samples from the island of Vis, as described previously. The ratio between the circulating levels of Aβ42 and Aβ40 was calculated after exclusion of outlier measurements as described in Section 2.5.4. Plasma Aβ42:Aβ40 ratios have been shown to be a useful biomarker for identifying individuals at increased risk of cognitive impairment and LOAD (Graff-Radford *et al*., 2007). Low Aβ42:Aβ40 ratios have been suggested to indicate decrease in the concentration of Aβ42 peptide in plasma as a result of increased deposition of Aβ42 in the brain (van Oijen *et al*., 2006; Graff-Radford *et al*., 2007).

A genome-wide QTL association analysis of the Aβ42:Aβ40 ratio in plasma was carried out with the aim of identifying genomic loci that potentially influence Aβ42 deposition in the brain. The 308,140 SNPs from the Illumina Hap300 array were used in the final analysis.

The genome-wide association analysis of the Aβ42:Aβ40 ratio revealed 6 SNPs which passed the threshold for suggestive significance ($P \leq$ 1.0E-05). These are located in chromosomal regions 1q41, 1q44, 1p13, 2q22, 10q23 and 15q23. Five other loci were detected which showed potential association in chromosomal regions 1q44, 6q24, 8p23, 19q13 and 22q13 (**Figure 5.6**).

The marker showing the highest association with plasma Aβ42:Aβ40 ratio was SNP, rs2886199, in intron 40 of the Usher Syndrome 2A (*USH2A*) gene in chromosomal region 1q41 (P = 1.1E-06). SNP rs2886199 explained about 2.7% of the Aβ42:Aβ40 ratio variance and the minor allele was associated with a reduction in the ratio (β = -0.39 ± 0.08 SE) (**Table 5.3**). *USH2A* rs2886199 lies near a region of high conservation and "regulatory potential" (UCSC Human Genome Browser, March 2006 assembly) about 20 kb from another associated SNP, rs11120605 (β = -0.206, P = 1.1E-03) in the same intron.

Another marker showing suggestive association on chromosome 1 was rs755854, located in chromosomal band 1p13 (P = 1.5E-06) potentially explaining 2.7% of the phenotypic variance (**Table 5.3**). This SNP is in intron 6 of the Immunoglobulin Superfamily member 2 (*IGSF2*) gene in a region of high "regulatory potential" (King *et al*., 2005; Kolbe *et al*., 2004; UCSC Human Genome Browser, March 2006 assembly). However no other potential SNPs within the same region satisfied the criteria for suggestive or potential associations.

SNP rs13379548 in chromosomal region 15q23 passed the suggestive association threshold (P = 2.1E-06) and potentially explained 2.6% of the Aβ42:Aβ40 ratio variance (**Table 5.3**). The minor allele of this SNP, which is located 120 kb upstream of a hypothetical gene, *FLJ13710*, was found to be associated with a decrease in the Aβ42:Aβ40 ratio in plasma (β = -0.036 ± 0.08 SE).

A peak showing suggestive association with plasma Aβ42:Aβ40 ratio was observed at a 2q22 locus. This peak spanned about 375 kb of the 686.8 kb Thrombospondin, Type I, Domain-containing 7B (*THSD7B*) gene (**Table 5.3**).

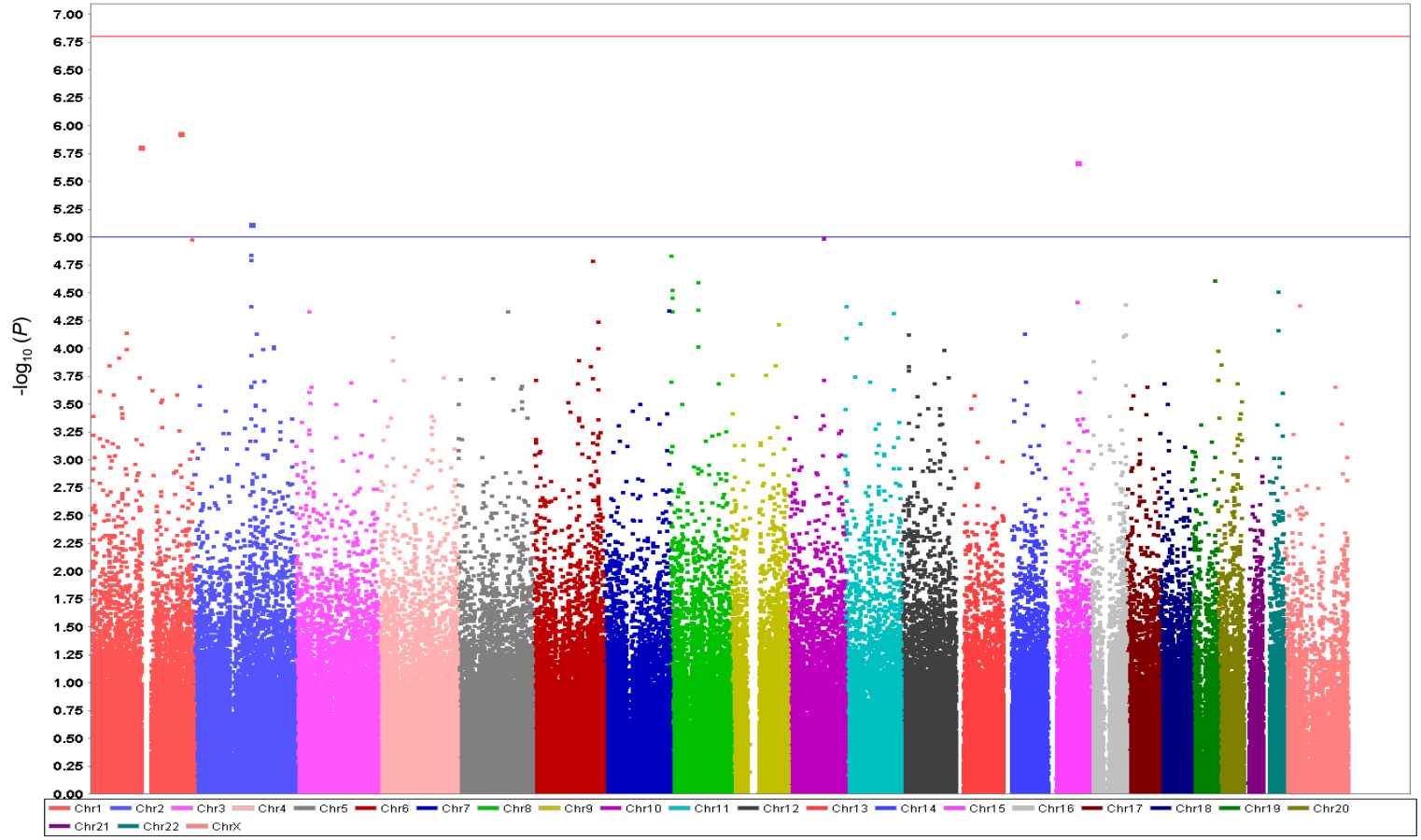


Figure 5.6 Results of the genome-wide quantitative trait locus association analysis of plasma Aβ42:Aβ40 ratio in the Vis population.

Results of the genome-wide QT association scan using 308,140 SNPs are shown for each chromosome. Analysis was carried out using normalised levels of Aβ42:Aβ40 measurements in 966 individuals from the island of Vis, after correcting for age and relatedness. Unadjusted $\log_{10}$ transformed *P*-values for association are presented on the y-axis. The threshold for genome-wide significance (red line) was set using a Bonferroni correction. Suggestive significance is marked by the blue line. Figure obtained from Haploview v. 4.0 (Barrett *et al*., 2005).

SNP	Minor allele	Chromosome	Position (bp)	MAF	N_{chrom}	$P_{association}$	R^2	P_{HWD}	β	S.E.	Genomic context
rs2886199	C	1q41	212274233	0.116	1920	1.1E-06	0.027	0.875	-0.39	0.08	*USH2A* intron 40
rs755854	C	1p13	117275624	0.259	1904	1.5E-06	0.027	0.175	0.29	0.06	*IGSF2* intron 6
rs13379548	C	15q23	69689719	0.123	1932	2.1E-06	0.026	1.000	-0.36	0.08	120kb upstream of *FLJ13710*
rs1349159	G	2q22	138069011	0.495	1928	7.5E-06	0.023	0.092	0.22	0.05	*THSD7B* intron 11
rs3904537	T	10q23	85435487	0.145	1924	1.0E-05	0.022	0.516	0.32	0.07	Gene desert
rs10924216	C	1q44	241989560	0.412	1878	1.0E-05	0.023	0.589	-0.23	0.05	Gene desert
rs1532583	T	8p23	3336092	0.292	1926	1.4E-05	0.022	0.532	-0.24	0.06	*CSMD1* intron 11
rs1038390	G	6q24	143868795	0.224	1932	1.6E-05	0.021	0.642	0.27	0.06	*FUCA2* intron 2
rs1654537	C	19q13	56158328	0.297	1896	2.4E-05	0.021	0.137	-0.24	0.06	*KLK6* intron 4
rs5756540	A	22q13	35906705	0.184	1928	3.0E-05	0.020	0.589	-0.27	0.06	C1QTNF6 intron 1
rs4896663	G	6q24	143874819	0.389	1932	1.8E-04	0.016	0.375	0.19	0.05	3kb upstream of *FUCA2*

Table 5.3 Summary of the genome-wide QTL association analysis of the plasma Aβ42:Aβ40 ratio in the Vis population.

Details of the most significant associations from the genome-wide association analysis of the plasma Aβ42:Aβ40 ratio in 966 individuals from the island of Vis. The analysis was carried out using an additive model of the minor allele after correcting for age and relatedness. Markers showing suggestive significant associations ($P \leq$ 1.0E-05, shown above the dotted line) and those showing potential associations (shown below the dotted line). Minor allele frequencies (MAF) and the number of chromosomes (N) in the population sample are also shown. R^2 represents the proportion of the phenotypic variance explained by each SNP. The direction of effect exerted by the minor allele of each SNP is represented by β, the regression coefficient and its standard error (S.E.). Values of β are arbitrary as these are based on the transformed distributions of Aβ42:Aβ40 in plasma. All markers were found to be in Hardy-Weinberg equilibrium. The analysis was carried out using the software package PLINK (Purcell *et al.*, 2007).

The most significantly associated SNP at this locus was rs1349159 (P = 7.5E-06), in intron 11 of *THSD7B*, potentially explaining 2.3% of the phenotypic variance. The rs1349159 minor allele was associated with a significant increase in the ratio of Aβ42:Aβ40 in plasma (β = 0.22 ± 0.05 SE). The plasma Aβ42:Aβ40 ratio in addition to SNP rs1349159, was found to be associated with seven other SNPs within *THSD7B* gene and detailed analysis revealed high levels of pairwise LD between these markers in the Vis data set (**Figure 5.7**). These intronic SNPs covered a region of about 300 kb between exons 8 and 17 near the 3′ end of the *THSD7B* gene.

An association peak was observed at 10q23 which reached suggestive levels of significance for rs3904537 (P = 1.0E-05), potentially accounting for 2.2% of the phenotypic variance (**Table 5.3**). This SNP is located in a "gene desert" defined as having no known genes within a 1Mb distance on either side of the marker.

Another suggestive association was found between plasma Aβ42:Aβ40 ratio and SNP rs102924216 (P = 1.0E-05), located in chromosomal region 1q44. This SNP is also located in a gene desert with no known genes in its vicinity.

A SNP in chromosomal region 8p23, rs1532583, showed potential genome-wide association significance (P = 1.4E-05) and was found to be situated in intron 1 of the CUB and Sushi Multiple Domains 1 (*CSMD1*) gene (**Table 5.3**). This SNP potentially explained 2.2% of the phenotypic variance and the minor allele resulted in a decrease in the plasma Aβ42:Aβ40 ratio (β = -0.24 ± 0.06 SE). Further analysis of this association peak revealed another cluster of SNPs within the same 2.05 Mb gene which reached significance levels of $P \leq$ 1.0E-04, located 1.1 Mb away from rs1532583 (**Figure 5.6**). Analysis of LD at this locus did not reveal any particular patterns around the two association peaks. Furthermore, no reports of predicted regulatory potential at these sites have been documented.

A potential association peak in chromosomal region 6q24 identified two SNPs (rs1038390 and rs4896663) associated with an increase in the Aβ42:Aβ40 ratio in plasma (**Table 5.3**). The marker more significantly associated with the trait, rs1038390 (P = 1.6E-05), resides in intron 2 of the Fucosidase, Alpha-L- 2 (*FUCA2*) gene. The other SNP, rs4896663 ($P_{association}$=1.8E-04) was found to be located 3 kb upstream of the same gene. *FUCA2* encodes a plasma alpha-L-fucosidase precursor protein. However, the validity of this finding would need to be replicated as the

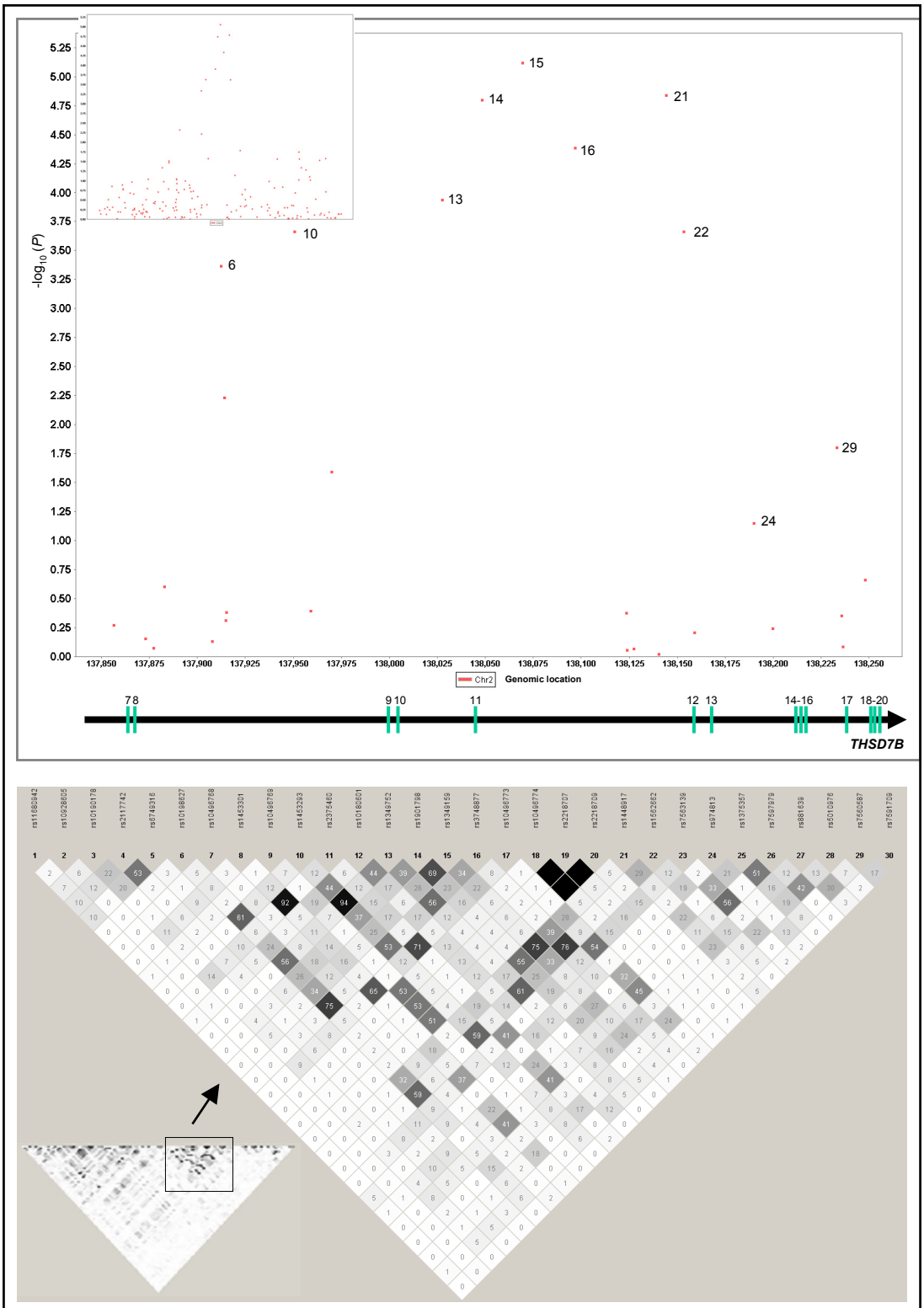


Figure 5.7 Analysis of the suggestive association peak at 2q22 in relation to the Aβ42:Aβ40 ratio in plasma.

Association of the Aβ42:Aβ40 ratio with a 2q22 locus spanning the 3′ end of the *THSD7B* gene (drawn below the association figure, to scale). Levels of pairwise linkage disequilibrium (LD) terms of the squared correlation (r^2) are also presented. Decreasing shades of black represent reductions in the level of pairwise LD. Markers which satisfied $P_{association} \leq$ 1.0E-04 (labelled in the association figure) all showed strong pairwise LD ($r^2 > 0.50$). The analysis was carried out in 492 unrelated individuals from Vis island, using Haploview. Vertical bars mark the last 14 exons of *THSD7B* (exons 7-20). The direction of transcription is shown by the black arrow head.

association peak consists only of the two SNPs mentioned above. Similarly the SNP, rs5756540 ($P_{association}$ = 1.3.0E-05), together with rs4821599 at a 22q13 locus showed potential associations with the Aβ42:Aβ40 ratio in plasma (**Table 5.3**). SNP rs5756540 is located in the first intron of the C1q and Tumour Necrosis Factor-related protein 6 (*C1QTNF6*) gene, encoding a secreted protein involved in the activation of the complement pathway and SNP rs4821599 is located approximately 20 kb downstream of *C1QTNF6* SNP rs5756540. This region was not assessed further since it did not satisfy the criteria for potential or suggestive genome-wide association.

Finally a potential association was detected with SNPs in chromosomal region 19q13 which contains a cluster of 15 Kallikrein genes (*KLK*), encoding a family of serine proteases (**Figure 5.6**). The highest associated marker, rs1654537 (*P* = 2.4E-05) is located in intron 4 of the *KLK6* gene and the minor allele is associated with a significant decrease in the Aβ42:Aβ40 ratio in plasma (β = -0.24 ± 0.06 SE), potentially explaining 2.1% of the phenotypic variance (**Table 5.3**). SNP rs1654537 is localised to a highly conserved region about 50 bp downstream of *KLK6* exon 4, exon-intron junction. This locus was not analysed further since no other SNPs showed suggestive or potential genome-wide association with plasma Aβ42:Aβ40 ratio.

5.3 Summary and discussion

In this chapter the results of a genome-wide association analysis of plasma Aβ40, Aβ42 and the Aβ42:Aβ40 ratio were described. Association analyses were carried out using the Illumina Human Hap300 Array consisting of 317,503 SNPs, after adjusting for covariate effects and relatedness among subjects. The aim was to identify genetic variation correlated with changes in the concentrations of Aβ species in plasma, potentially influencing susceptibility to LOAD. Further research into the mechanism of action of candidate associated loci that are replicated in additional/independent samples will help to identify new pathways involved in AD pathogenesis.

Genome-wide association analysis of plasma Aβ40

Whole genome association analysis of plasma Aβ40 revealed a number of loci which reached the genome-wide threshold for suggestive significance ($P \leq$ 1.0E-05) after correcting for multiple testing using a stringent Bonferroni correction (**Figure 5.1**). However, none of the associated SNPs showed full genome-wide significance after correction for multiple testing, so replication of these results is required. The putative loci modifying Aβ40 were identified in chromosomal regions 8p, 8q, 9p, 9q, 14q, 20p and Xp (**Figure 5.1**, **Table 5.1**).

Two of the most significant associations were observed in the *ASTN2* gene at 9q33. *ASTN2* rs2900131 (P = 2.1E-06) and rs7866525 (P = 3.2E-06) minor alleles resulted in a significant decrease in Aβ40 concentrations using an additive model ($\beta_{rs2900131}$ = -0.23 ± 0.05 SE, $\beta_{rs7866525}$ = -0.22 ± 0.05 SE). Further analysis revealed a cluster of SNPs near the transcription start site of the *ASTN2* alternative transcript D with *P*-values for association with plasma Aβ40 less than 1.0E-03. *ASTN2* encodes a transmembrane protein which is highly expressed in post-mitotic neuronal precursors in the cerebellum, hippocampus and cerebrum of the developing brain (Zheng *et al.*, 1996). *In vitro* studies have shown that astrotactin acts as a ligand for neuron-glia binding during neuronal migration and is likely to play a role in the movement of neurons along glia fibres (Hatten & Mason, 1990; Zheng *et al*., 1996). Furthermore, astrotactin is shown to be missing or defective on granule cells from the neurological mutant mouse model that suffers a failure of glial-guided neuronal migration (Edmondson *et al*., 1988). Therefore genetic variation in this gene may influence transcription of one or more isoforms during the development of the brain.

Other genomic loci which were found to show suggestive association with changes in plasma Aβ40 concentrations were the *DMD* locus on chromosome Xp21 and a number of other SNPs lying within gene deserts (**Table 5.1**). Some of the SNPs, including *DMD* rs12009327, were found to be located in regions of high conservation and strong predicted regulatory potential, indicating a potential role in regulation of gene expression. The percentage of the phenotypic variance explained by the associated markers was in the range of 2.0-2.4%. The *DMD* gene is highly complex, containing at least eight independent, tissue-specific promoters including a

brain-specific promoter. DMD mRNA is differentially spliced, producing a range of different transcripts which encode a large set of protein isoforms some of which are expressed in brain astrocytes (Nicchia *et al*., 2008). The dystrophin protein is a large, rod-like cytoskeletal protein, which is found at the inner surface of muscle fibers. Dystrophin is part of the dystrophin-glycoprotein complex (DGC), which bridges the inner cytoskeleton and the extracellular matrix and is mutated in Duchenne muscular dystrophy (Lindenbaum *et al*., 1979).

In summary, while a number of new and interesting candidate genes were identified, there was no overlap with linkage peaks identified for Aβ40. All suggestive or potential associations now require independent replication.

Genome-wide association analysis of plasma Aβ42

Testing for association of plasma Aβ42 concentration resulted in the identification of a number of loci which passed the threshold for suggestive significance. Potential associations were observed with SNPs in chromosomal regions 2q32, 5q14, 9q33, 15q21, 20q13 and Xp11 (**Figure 5.4**). Some of these SNPs were within genes, including *TMEFF2* (2q32), *ATP8B4* (15q21) and *PREX1* (20q13). Two polymorphisms (rs1113932 and rs4267464) were identified in intron 2 of the *TMEFF2* gene ($P_{association} \leq$ 1.0E-05) which showed suggestive associations with plasma Aβ42 concentrations (**Table 5.2**). *TMEFF2* encodes a transmembrane protein, predominantly expressed in the brain that acts as a survival factor for hippocampal and mesencephalic neurons. The TMEFF2 protein has been found in AD plaques (Siegel *et al*., 2006) and genetic variation within the gene may contribute to Aβ42 deposition. Three SNPs in the peak of association located in intron 2 of the *TMEFF2* gene were found to be highly correlated both with each other and with Aβ42 concentration through strong levels of pairwise LD (**Figure 5.5**). The strength of association falls in intron 4 of the gene *TMEFF2* where a region of "recombination hotspots" separates the associated SNPs (in a region of high LD) from the surrounding region (**Figure 5.5**). The presence of a region with high LD separated by recombination from its surrounding regions will facilitate the identification of potentially functional polymorphisms using LD mapping (Goldstein, 2001). The exact mechanism by which the function or expression of TMEFF2 is

affected by genetic variation in and around *TMEFF2* requires further analysis. The major objective at this stage is to replicate the association results in an independent population sample or in LOAD case-control samples.

Potential associations were detected with plasma Aβ42 concentration in the *ATP8B4* (rs2067885) and *PREX1* (rs6122718 and rs6095239) genes (**Table 5.2**). Association studies have previously identified the 15q21 locus, containing the *ATP8B4* gene, as a potential LOAD susceptibility locus (Li *et al*., 2008). The *ATP8B4* SNP rs2067885 is located in a region of high sequence conservation in intron 2, which may indicate involvement of a region regulating gene expression. The PREX1 protein is found mainly in the cytoplasm and is activated by phosphatidylinositol-3,4,5-trisphosphate and the beta-gamma subunits of heterotrimeric G proteins. PREX1 has been shown to be involved in neurotrophin-derived signalling and neuronal migration during development in mice (Yoshizawa *et al*., 2005), similar to astrotactin 2.

The Aβ42 association analysis found other potential associations with SNPs of low minor allele frequency in the Vis data set (**Table 5.2**). However, the probability of obtaining false-positive results in association tests is increased for SNPs of low frequency, so these results in particular need to be replicated. The proportion of the Aβ42 phenotypic variance accounted for by these potential associations ranged from 1.8% to 2.8% (**Table 5.2**).

Genome-wide association analysis of plasma Aβ42:Aβ40 ratio

Genome-wide association analysis of the ratio of Aβ42:Aβ40 in plasma revealed a number of loci showing suggestive associations, in chromosomal regions 1q41, 1q44, 1p13, 2q22, 6q24, 8p23, 10q23, 15q23, 19q13 and 22q13 (**Figure 5.6**). Lower ratios of Aβ42:Aβ40, accompanied by high Aβ40 concentration, have been associated with increased risk of AD (Graff-Radford *et al*., 2007). This could indicate a higher proportion of Aβ42 being deposited in the brain early in the Alzheimer disease process due to the presence of a risk allele.

Variants associated with reductions in the plasma Aβ42:Aβ40 ratio included the following: *USH2A* SNP rs2886199 is located in intron 40 of this gene which is

associated with mutations in Usher's syndrome type 2 and autosomal recessive retinitis pigmentosa (Eudy *et al*., 1998; Liu *et al*., 1999). The *USH2A* protein is a basement membrane-associated protein which may be important in the development and homeostasis of the inner ear and retina. It contains laminin EGF motifs, a pentraxin domain, and many fibronectin type III motifs.

CSMD1 SNP rs1532583 is located in intron 11 of the gene and potentially explains 2.2% of the phenotypic variance (**Table 5.3**). *CSMD1* is a regulator of the classical pathway of complement activation which is highly expressed in the central nervous system. However, analysis of LD in the region did not reveal any associated functional or coding variants and the associated SNPs were not located in regions of high sequence conservation or regulatory potential. The validity of all these findings needs to be replicated and examined further using a higher resolution genetic analysis of the region.

A SNP in intron 11 of the *THSD7B* gene reached suggestive genome-wide significance for association with the ratio of Aβ42:Aβ40 in plasma (**Table 5.3**). *THSD7B* encodes a thrombospondin type-1 domain-containing secreted protein with a catalytic domain which cleaves large aggregating chondroitin sulphate proteoglycans (aggrecan and versican). Chondroitin sulphate proteoglycans have been found in senile plaques of human AD tissue and have been shown to inhibit glial growth in AD brains (Canning *et al*., 1993). Chondroitin sulphate proteoglycans have also been shown to interact with Aβ species in the blood vessels of the brain (Snow *et al*., 1995). Detailed analysis of the locus containing the *THSD7B* gene showed a cluster of associated SNPs all of which are intronic, spanning a region of about 300 kb, all in strong LD with each other ($r^2 > 0.60$) (**Figure 5.7**).

Another potentially interesting candidate for association with plasma Aβ42:Aβ40 ratio was identified as the *KLK6* gene. The marker showing the strongest association, rs1654537, was found to be located in a highly conserved region about 50 bp downstream of an exon-intron junction which could influence gene expression through splicing. The KLK6 protein has been immunolocalised in AD brains and, in tissue culture, the enzyme has been found to generate amyloidogenic fragments from the amyloid precursor protein, suggesting potential

involvement in Alzheimer's disease (Little *et al*., 1997). Furthermore, the brains of LOAD patients have been found to contain significantly lower concentrations of KLK6 than normal (Zarghooni *et al*., 2002) indicating a protective role for this serine protease. The same pattern has been reported in LOAD CSF samples (Mitsui *et al*., 2002). Therefore genetic variants in *KLK6* could influence its protease activity or expression level, resulting in higher levels of Aβ42 deposition in the brain and changes in the plasma Aβ42:Aβ40 ratio.

Variants in the amyloid precursor protein (*APP*) gene and the enzymes involved in APP processing at the cellular level have been shown to be associated with autosomal dominant familial AD (St George-Hyslop *et al*., 1987). In the Vis data set, no association was observed between concentrations of Aβ species in plasma and the variants of *APP* (chromosome 21), *PSEN1* (chromosome 14), *PSEN2* (chromosome 1) or *APH1A* (chromosome 1) components of γ-secretase after screening for associations in the regions harbouring those genes. However, as **Table 5.4** shows, the number of SNPs in the Illumina Hap300 array, which cover these genes is low (*APP* (290 kb), 40 SNPs; *PSEN1* (83.5kb), 4 SNPs; *PSEN2* (24.8 kb), 2 SNPs; *APH1A* (3.5 kb), no SNPs). Therefore, fine-mapping of the genomic regions encompassing the genes involved in direct processing and production of Aβ species may identify stronger associations.

In conclusion, genome-wide association analysis of plasma Aβ40, Aβ42 and the ratio of Aβ42:Aβ40 was carried out for the first time in a general population sample and revealed biologically interesting candidate loci potentially influencing susceptibility to AD. A number of hits reached suggestive levels of significance without any signs of association in the neighbouring SNPs. The results of this association analysis therefore need to be validated by means of replication in other data sets such as a second Croatian sample ($N \sim 1{,}000$) who have also recently been genome scanned using the Illumina Hap300 array or the LOAD cohorts of cases and controls available in the Younkin lab. This has become possible as part of a collaboration with the Mayo Clinic (Jacksonville, USA) aimed at the identification of genomic loci affecting predisposition to AD. With access to one of the world's most extensive collections of LOAD patients and controls, the findings of the current association analysis could be readily validated.

Gene	Genomic context	Start	Stop	Illumina SNPs in the gene (*n*)	Highest association with Aβ40, Aβ42, Aβ42:Aβ40 ratio (*P*)
APP	21q21	26,174,732	26,465,002	40	0.052, 0.064, 0.012
PSEN1	14q24	72,673,277	72,756,861	4	0.580, 0.370, 0.215
PSEN2	1q42	223,365,694	223,390,531	2	0.266, 0.488, 0.711
APH1A	1q21	147,050,878	147,054,383	-	-

Table 5.4 SNP coverage of the genes involved in familial Alzheimer disease by the Illumina Hap300 array.

The number of SNPs present in the Illumina human Hap300 array covering the genes involved in the familial autosomal dominant Alzheimer disease are shown together with the highest association observed in the genomic region of those genes. The genes analysed consisted of the amyloid precursor protein (*APP*) gene and genes encoding the three components of the γ-secretase protein involved in the cleavage of APP which results in the production of Aβ species. *P*-values show the significance of the highest association SNP in each gene. Positions are based on NCBI Build 35 assembly.

The findings of the current genetic analysis of plasma Aβ identified common genetic variation of relatively small effect size. It has been suggested that there are at least four other genetic loci with variants of moderate to large effect size, similar to the *APOE*E4* in Alzheimer disease (Daw *et al*., 2000). As shown in Chapter 3, there is sufficient power in the Vis population data set to identify variants of small effect by genetic association provided there is a strong LD relationship between the marker and the functional variant. Since a number of potentially relevant polymorphisms identified here failed to reach the full genome-wide significance threshold, increasing the sample size may help to detect variants of small effect size in this population. This will be possible using a second Croatian sample ($N \sim 1{,}000$) in whom a Human Hap300 genome-wide scan will soon be completed and plasma Aβ analysis is planned.

CHAPTER 6

BIOCHEMICAL AND LINKAGE ANALYSES OF PLASMA COMPLEMENT FACTOR H

6.1 Introduction

In this chapter the results of biochemical analysis of human plasma CFH have been described and the effects of a number of covariates on CFH plasma levels assessed. The quantitative variation observed in plasma CFH concentration is believed to be the result of combined genetic and environmental factors. It has been reported that 63% of the variation in plasma CFH concentration is genetically determined in a general population of Spanish origin ($h^2 = 0.63 \pm 0.07$; $P < 0.0001$) (Esparza-Gordillo *et al*., 2004). Furthermore, in a whole genome QT linkage analysis, Esparza-Gordillo and colleagues observed suggestive linkage to three genomic regions, 1q32, 2p21 and 15q22.

Common variants of the *CFH* gene have been implicated in age-related macular degeneration with rs1061170 (Y402H) showing strong association with AMD (Haines *et al*., 2005; Edwards *et al*., 2005; Klein *et al*., 2005; Zareparsi *et al*., 2005a). This polymorphism is located in a region of CFH that definitely binds a heparin-like glycosaminoglycan (GAG) (Prosser *et al*., 2007) and possibly binds C-reactive protein (CRP) (Pepys, 2008). Strong evidence from crystal structure analysis shows that GAG binding is altered by this amino acid substitution (Prosser *et al.*, 2007). Plasma CFH concentration was therefore investigated as a possible surrogate biomarker for age-related macular degeneration.

The aim was to determine concentrations of CFH in the plasma of individuals from two isolate populations of Croatian and Dutch origin, and to assess the effect of covariates, including *CFH* Y402H, on plasma CFH concentrations. Plasma CFH measurements will also be used as the quantitative trait in a genome-wide linkage analyses of markers associated with plasma CFH concentration in Vis.

Measurements of plasma CFH were also carried out in a Scottish series of AMD cases and controls to assess whether they significantly differed between the two groups hence assessing its potential as a biomarker for AMD. A sensitive "sandwich" ELISA was first developed (**Figure 6.1**) and optimised using purified full-length human CFH as the standard. A mouse monoclonal anti-human CFH antibody (OX23) was used for detection.

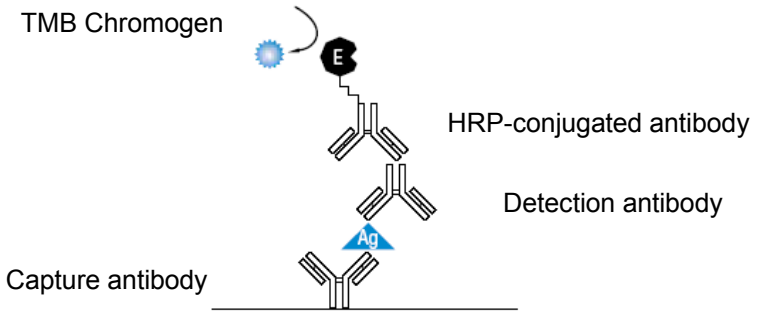


Figure 6.1 Diagrammatic presentation of "sandwich" ELISA of CFH.

A sheep polyclonal antibody was used as the capture antibody to pull down as much of the antigen as possible. A mouse monoclonal antibody was used as the detection antibody to provide improved specificity. Fluorometric detection was carried out at 450 nm.

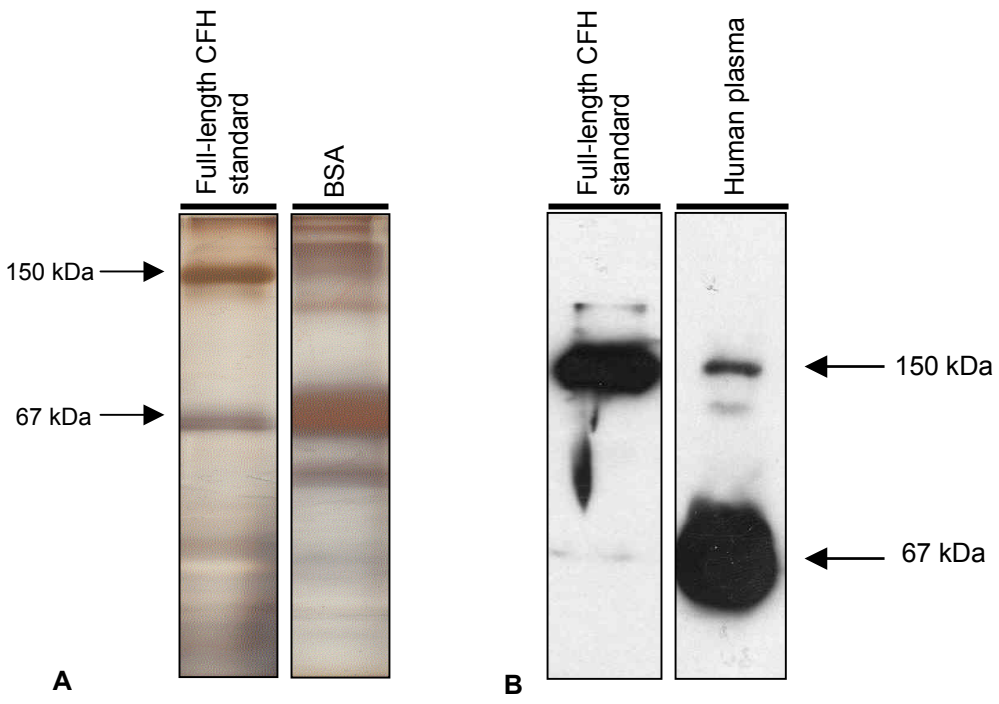


Figure 6.2 Testing the purity of the CFH standard and the specificity of the OX23 antibody.

A. Purity of the CFH standard was tested by silver staining a 10% SDS-polyacrylamide gel containing 18.75 μg of the 150 kDa purified full-length CFH (left) compared with a 67 kDa 98% purified lyophilised BSA sample (right) which confirmed the presence of serum albumin in the CFH standard. **B.** Western blot analysis of 7.5 μg of purified CFH (left) and human plasma (right). CFH was detected using 13.6 ng of the monoclonal OX23 primary antibody and 1.16 μg of HRP-conjugated goat anti-mouse secondary antibody.

6.2 Testing the purity of CFH standard and specificity of OX23 antibody

The purity of the external protein standard can cause variation in measurements between assays. The purity of the full-length human CFH (kindly provided by Dr. R. B. Sim, MRC Immunochemistry Unit, Oxford) was therefore tested using a highly sensitive silver stain (**Figure 6.2-A**) as described in Chapter 2. The purified full-length CFH was found to be of relatively high purity (estimated as about 90% from the silver stained gel). The presence of BSA in the purified CFH standard, confirmed by the silver stained BSA sample (**Figure 6.2-A**), is believed to enhance the stability of the protein. The concentration of the full length CFH was measured previously (Dr. R. B. Sim) and used as the concentration of the standard.

The specificity of the mouse monoclonal anti-human CFH (OX23) antibody (kindly provided by Dr. R. B. Sim) for binding to plasma CFH was also examined in a western blot using the full-length human CFH protein standard as a positive control (**Figure 6.2-B**). Detection of the full-length CFH protein by the OX23 antibody was confirmed in a western blot analysis (**Figure 6.2-B**) using the purified human CFH standard. Western blot analysis of human plasma showed that the antibody recognised the full-length human CFH protein (150 kDa) but cross-reacted strongly with serum albumin at high concentrations in the human plasma (**Figure 6.2-B**). This suggested that the antibody had been raised to a BSA-conjugated fragment of CFH. However, detection of albumin would depend on the polyclonal capture antibody reacting with serum albumin. The lack of detection of albumin by the polyclonal capture antibody was confirmed by the absence of signal from the "blank" wells in CFH ELISA plates, which were coated with the capture antibody and blocked with BSA.

6.3 CFH assay optimisation

Assay optimisation was carried out in order to determine the optimal conditions for a number of factors including the concentration of the antibodies, the concentration of the CFH standard, the sample dilution factor and the incubation times.

6.3.1 Optimisation of antibody and standard concentrations

The optimal concentration for the detection antibody to provide adequate assay sensitivity was found to be 85 ng/ml, since the curve for optimising the concentration of this antibody measured a wide range of absorbances corresponding to 1 μg/ml of CFH standard at four different dilutions of the HRP-conjugated antibody (**Figure 6.3-A**, light blue line). The optimal concentration for the HRP-conjugated antibody was found to be 145 ng/ml, since the absorbance value was sufficiently high (A_{450nm} = 2.5) at this dilution of antibody to provide adequate assay sensitivity (**Figure 6.3-A**, vertical green line). **Figure 6.3-B** shows the effect of two differing concentrations of the capture antibody on the level of absorbance detected. Both concentrations showed optimal coverage in the range of 9.4-300 ng/ml of CFH standard. A concentration of 1.2 μg/ml of capture antibody was selected as optimal since it differed only negligibly from the 2.4 μg/ml concentration.

6.3.2 Optimisation of sample dilutions

Dilutions of plasma extracted from blood samples donated by local volunteers were used in duplicate to determine the optimal plasma dilution factor, as shown in **Figure 6.3-C**. The optimal sample dilution factor was found to be 1 in 5,000, assuming an approximate CFH concentration range of 100-600 μg/ml (Esparza-Gordillo *et al*., 2004). The aim was to ensure that plasma samples containing CFH in the range 100-600 μg/ml would be diluted to within the range of purified CFH standard concentrations used to construct the standard curve.

6.3.3 Optimisation of incubation time with TMB chromogen

The final stage of the assay is an enzymatic reaction in which tetramethylbenzidine (TMB) chromogen is oxidized by horseradish peroxidase in the presence of hydrogen peroxide to yield a chinoid structure of intense blue colour. The reaction is affected by a number of factors, such as temperature, time and pH. The effect of temperature could not be assessed due to lack of a temperature-controlled instrument. The effect of varying the time course of incubation with chromogen is shown in **Figure 6.4**.

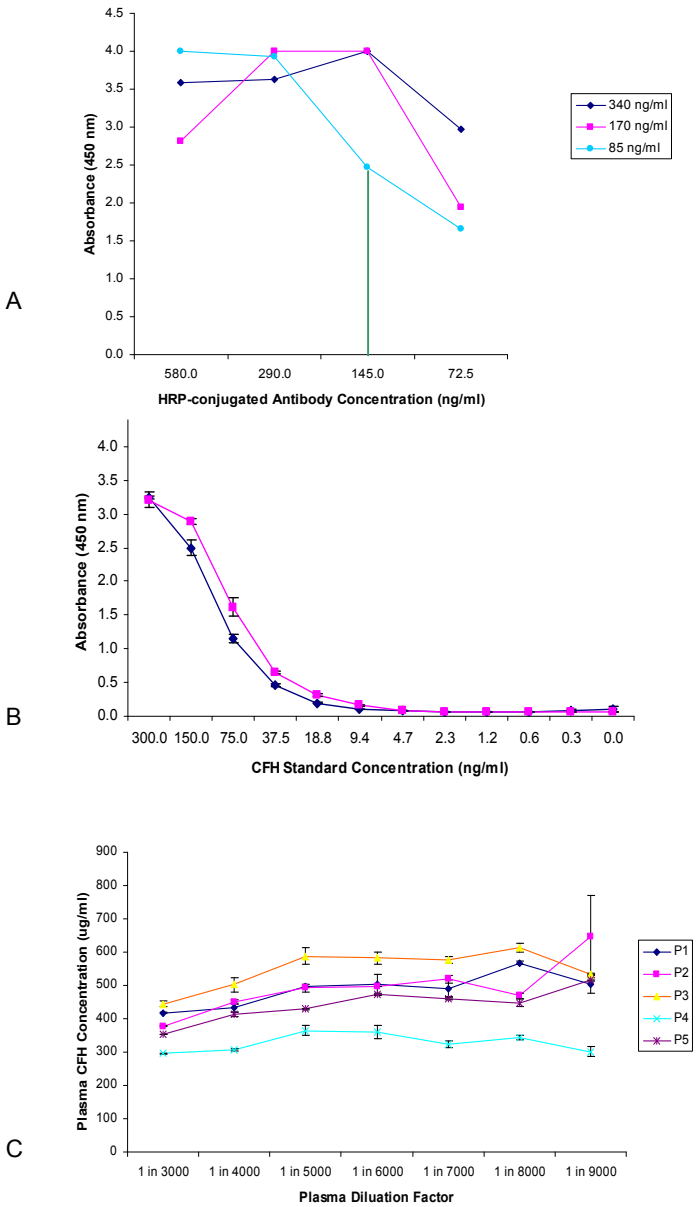


Figure 6.3 Titration curves showing the stages of CFH assay optimisation.

A. Coloured lines show the range of absorbance values covered by varying concentrations of the OX23 detection antibody. The optimal concentration of the OX23 detection antibody was found to be 85 ng/ml since this measured a wide range of absorbances corresponding to 1μg/ml of CFH at four different dilutions of the HRP-conjugated antibody. The vertical green line shows the optimal concentration of HRP-conjugated antibody resulting in sufficiently high absorbance level at 85 ng/ml of OX23. **B.** Coloured lines represent two different concentrations of the capture antibody (red: 2.4 μg/ml and blue: 1.2 μg/ml) over a range of CFH standard concentrations. 1.2 μg/ml was selected as the optimal concentration of the capture antibody since it differed negligibly from the 2.4 μg/ml concentration. **C.** Effect of varying dilutions of five plasma samples (labelled P1-P5), ranging from 1 in 3000 to 1 in 9000 dilutions, on detection of CFH concentration. Error bars represent standard error of the mean.

The optimal incubation time was found to be 60 minutes. The pH was kept at neutral, which is widely used for this reaction.

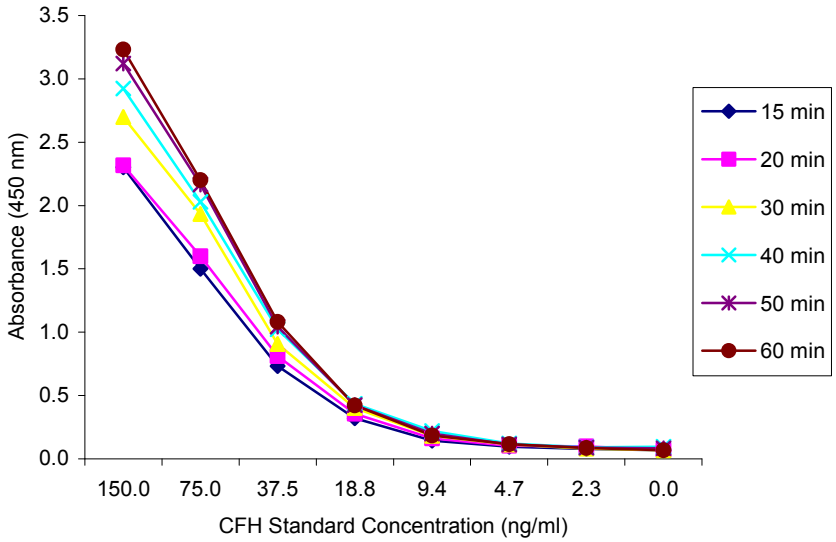


Figure 6.4 Effect of varying incubation times with TMB chromogen.

Optimal absorbance measurements for ELISA of CFH were observed after 60 minutes. The figure shows that at the 60 minutes incubation time the reaction has not yet entered a plateau phase indicating saturation of the peroxidase enzyme.

6.3.4 Analysis of intra-experimental variation

Five internal control plasma samples were used to assess intra-experimental variation between samples. **Figure 6.5-A** shows the variation observed between at least 8 replicates of each sample. Coefficients of variation (CV) were all below the 5% threshold, showing that there were no significant differences between sample replicates. The small variation observed could be accounted for by pipetting errors.

6.3.5 Analysis of inter-experimental variation

Inter-experimental variation was measured using the same internal controls. Comparison of the mean of three replicates of each sample from three different assays showed minimal variation (**Figure 6.5-B**).

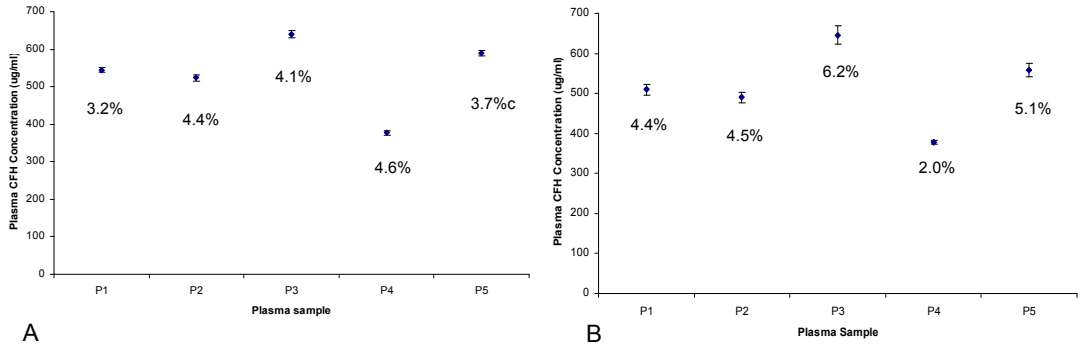


Figure 6.5 Analysis of inter and intra-experimental variation of CFH ELISA assay.

A. Intra-experimental variation within replicates of five plasma sample are shown (P1-P5). Each point represents the mean of at least 8 replicates. Variation was measured in terms of coefficient of variation (CV) shown below each point. **B.** Inter-experimental variation based on the mean of three replicates from three different assays are shown. Each point represents the mean of three measurements from three independent assays. The CVs are shown below each point. Error bars are standard error of the mean.

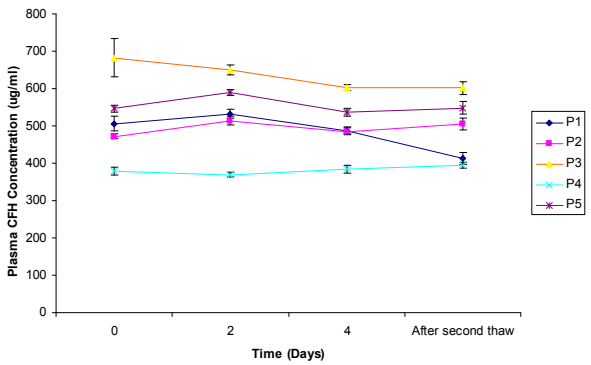


Figure 6.6 Analysis of variation after periods of sample freeze-thawing

Graph showing plasma CFH levels measured in triplicate for five samples at four time-points: day 0 (first thaw), day 2, day 4 and after a second thaw. Error bars represent standard error of the mean. No significant differences were observed.

	0	2	4	After second thaw	MEAN	SD	CV (%)	*P*
P1	506.36	532.66	487.85	422.47	487.34	46.99	9.64	0.099
P2	471.37	514.39	484.78	506.06	494.15	19.65	3.98	0.188
P3	682.52	650.54	603.26	602.38	634.68	39.03	6.15	0.176
P4	379.06	369.56	384.70	395.59	382.23	10.88	2.85	0.505
P5	546.08	589.59	536.88	548.34	555.22	23.44	4.22	0.878

Table 6.1 Assay variation observed between different time-points.

No statistically significant differences were observed between initial measurements (day 0) and after a second episode of thawing ($P > 0.05$). The table also demonstrates the stability of CFH measurements for samples which have been kept at 4^{o}C for 2 and 4 days post-thawing and refrozen. CV: coefficient of variation, SD: standard deviation, *P*-value is calculated for the two-tailed t-test of the differences between day 0 and the second thaw.

The slight increase in CV in samples P3 and P5 was believed to be due to pipetting errors and experimental variation between assays, for example the use of different batches of reagents.

6.3.6 Analysis of plasma CFH concentration after sample freeze-thawing

Repeated freeze-thawing can have an adverse effect on levels of detectable proteins as it can damage the native structure of the polypeptide. Five crude EDTA plasma samples were thawed from –80^{o}C slowly on ice and stored at 4^{o}C for four days. Thereafter, they were re-frozen at –80^{o}C overnight and re-thawed on the following day. Plasma CFH levels were measured at each time point (**Figure 6.6**). Analysis of the results showed no statistically significant differences ($P > 0.05$) between CFH measurements following a second thaw (**Table 6.1**). The high concentration of CFH in human plasma most likely accounted for the stability of the protein on repeated freeze-thawing.

These results were of considerable significance for the following reasons. Firstly, they demonstrated that reliable CFH concentration readings could be obtained from samples which have been frozen twice. Plasma samples collected from the island of Vis and used for CFH measurements had been thawed once for the analysis of plasma cortisol. Therefore, the assay was capable of providing reliable and accurate plasma CFH concentrations. Secondly, the small variation observed in plasma CFH between days 0 and 4 showed that samples can be stored at 4^{o}C for up to 4 days after thawing and used in multiple experiments.

6.3.7 Analysis of pipetting techniques in CFH assay

In order to select the optimal pipetting method for the CFH ELISA, three different pipetting techniques were assessed and their contribution to the variation observed between sample measurements was evaluated. These included multi-channel, digital and manual single pipettes. Pipettes had previously been calibrated by weighing of multiple pipettes using water samples.

Analysis of the mean variation observed between CFH concentrations in eight plasma samples indicated that single pipette measurements differed significantly from measurements made with multi-channel pipettes. The mean variation was also lower using single manual pipettes than when using digital pipettes (**Figure 6.7**). The results indicated that the least variation was observed when using the single manual pipette (CV = 5.7%) compared with the other two methods. This was therefore used as the pipetting method for the CFH ELISA.

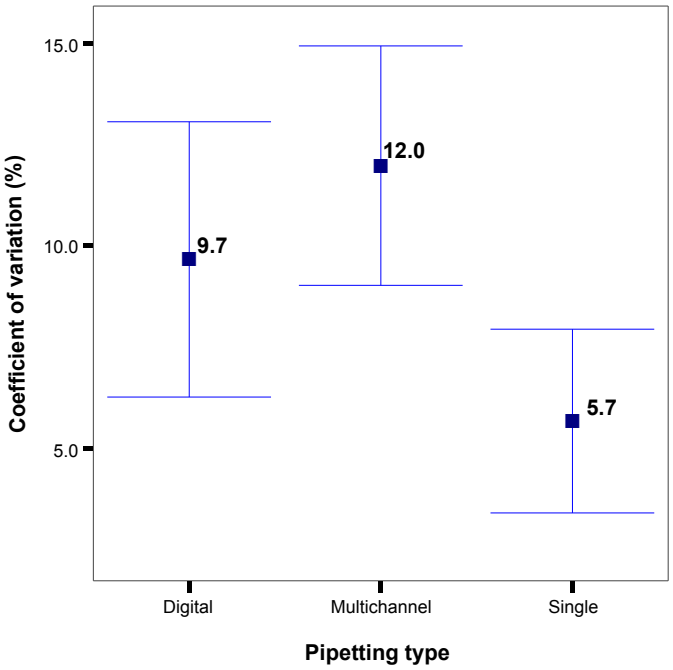


Figure 6.7 Effect of different pipetting techniques on assay variation

Diagram showing the coefficients of variation for different pipetting methods. The mean coefficient of variation is presented for each technique. Values are calculated from duplicates and are based on eight sample measurements. The single manual pipette, showing the least variation (CV=5.7%), was used as the pipetting method for the ELISA experiments. Error bars show 95% confidence interval of the mean.

6.4 Plasma CFH measurements

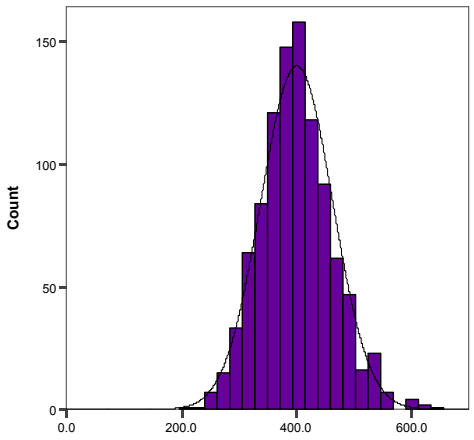
Plasma CFH concentrations were measured in 1,004 samples from the island of Vis (**Figure 6.8-A**), 490 samples from an isolated Dutch (Rucphen) population (**Figure 6.8-B**), 549 samples from a Scottish AMD series (**Figure 6.8-C**) and 88 samples from a Japanese AMD series (**Figure 6.8-D**) using the optimised assay, as described in Section 2.4.3.3. All measurements were carried out in triplicate and the means of sample replicates were then normalised for inter-assay variation using five internal controls. Details of the series analysed are presented in **Table 2.1**.

6.4.1 Analysis of plasma CFH in the island of Vis

The distribution of plasma CFH concentrations in samples collected from the isolated island of Vis showed a three-fold range of values, with a mean of 400.2 μg/ml. The distribution fitted a normal distribution with a kurtosis factor of 0.49 ± 0.15. Kurtosis has been defined as a measure of departure from normality (DeCarlo, 1997) and values below 0.80 refer to a normal distribution. The normality of the distribution reduces the probability of obtaining false positive results in future analyses.

6.4.1.1 Analysis of covariates

The association between covariates and plasma CFH concentration was assessed in a univariate analysis using a general linear model, with the statistical software SOLAR (Almasy & Blangero, 1998) which takes the relatedness of individuals into account by constructing a relationship matrix. The results of the covariate analysis are summarised in **Table 6.2**. SNP rs1061170 (*CFH* Y402H) has been shown to be strongly associated with AMD. The association between rs1061170 (*CFH* Y402H) and plasma CFH concentration was tested in an additive model based on the additive effect of the *CFH* rs1061170 minor allele C. Differences in mean effect of the minor allele on CFH concentration were estimated from β, the regression coefficient, in units of μg/ml.

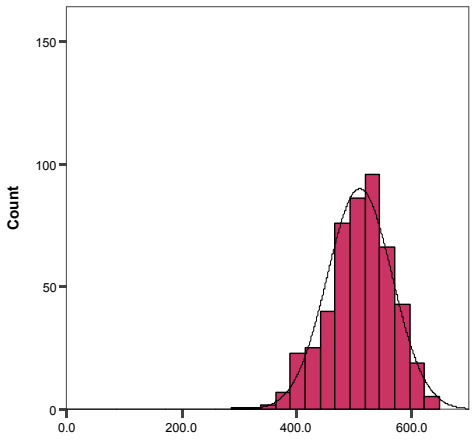


A

Plasma CFH Concentration (μg/ml)

Descriptive Statistics (Vis)

	N	Min.	Max.	Mean	S.D.	Kurtosis	
							S.E
Plasma CFH Concentration (ug/ml)	1004	218.0	654.1	400.2	62.4	.49	.15

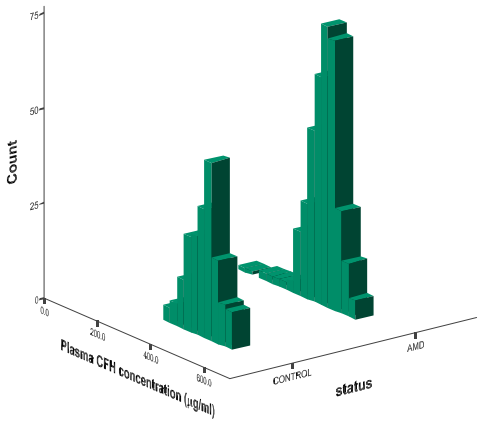


B

Plasma CFH Concentration (μg/ml)

Descriptive Statistics (Rotterdam)

	N	Min.	Max.	Mean	S.D.	Kurtosis	
							S.E.
Plasma CFH Concentration (ug/ml)	490	309.2	643.0	509.7	56.3	.11	.22



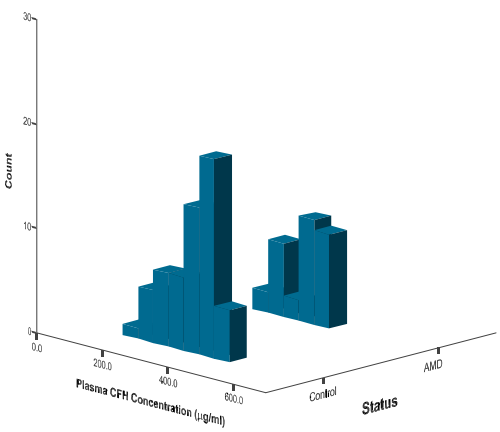
C

Descriptive Statistics (Scottish AMD)

	N	Min.	Max.	Mean	S.D.	Kurtosis	
							S.E.
Plasma CFH Concentration (ug/ml)	351	127.8	540.2	412.6	58.2	3.49	.26

Descriptive Statistics (Scottish Controls)

	N	Min.	Max.	Mean	S.D.	Kurtosis	
							S.E.
Plasma CFH Concentration (ug/ml)	198	302.2	541.6	428.3	51.9	-.32	.34



D

Descriptive Statistics (Japanese AMD)

	N	Min.	Max.	Mean	S.D.	Kurtosis	
							S.E.
Plasma CFH Concentration (ug/ml)	30	203.9	415.8	326.8	61.50	-1.092	.833

Descriptive Statistics (Japanese Controls)

	N	Min.	Max.	Mean	S.D.	Kurtosis	
							S.E.
Plasma CFH Concentration (ug/ml)	58	177.1	444.4	337.6	68.12	-.683	.618

Figure 6.8 Plasma CFH measurements in four study populations.

The distribution of unadjusted plasma CFH measurements in samples from the Croatian island of Vis **(A)**, the isolated Dutch (Rucphen) population **(B)**, a cohort of urban Scottish AMD cases and non-AMD controls **(C)** and a cohort of Japanese AMD cases and controls **(D)**. Population-specific descriptive statistics are presented below each graph. Kurtosis values smaller than 0.8 (Figures A and B) represent a normal distribution. Data were analysed using the software package SPSS v.12.

Covariate	χ^2	d.f.	*P*	β ± SE (μg/ml)	R^2
Age	1.69	1	0.193	n.a.	0.000
Sex	9.44	1	**0.002**	-11.2 ± 3.68	0.008
Age by Sex	8.60	1	**0.003**	0.9 ± 0.95	0.061
BMI	50.67	1	**1.1E-12**	3.3 ± 0.44	0.087
Smoking	0.22	1	0.642	n.a.	0.000
CFH Y402H	45.89	1	**1.2E-11**	-18.9 ± 2.78	0.048

Dependent variable: CFH (Mean 406 μg/ml ± 55.9 SD)

Table 6.2 Plasma CFH covariate analysis in Vis.

Results of the association analysis between plasma CFH and a number of covariates including age, sex, body mass index (BMI), *CFH* Y402H (rs1061170) (additive model of the minor allele) and interaction between sex and age. The analysis was carried out using 1,004 individuals from the island of Vis. *P*-values are presented for each covariate effect. β is the regression coefficient which indicates the change in the mean plasma CFH concentration (± standard error, SE). R^2 is the proportion of the unadjusted phenotypic variance explained. Analysis was carried out using a linear regression model implemented in the software package SOLAR (Almasy & Blangero, 1998)

SNP	MAF	Genotypic Frequency (%)	HWD *p*-value
rs1061170	0.39	TT (36.3), TC (49.1), CC (14.9)	0.50

Table 6.3 Descriptive statistics for rs1061170 (*CFH* Y402H) in the Vis population.

rs1061170 allele and genotypic frequencies were measured in 472 unrelated individuals from the island of Vis. Deviation from Hardy Weinberg equilibrium was also assessed. MAF: minor allele frequency, HWD *P*-value were calculated for deviation from Hardy Weinberg equilibrium. Analysis was carried out using the software package PLINK (Purcell *et al*., 2007)

The rs1061170 SNP was incorporated as a potential covariate into a SOLAR polygenic regression model that used plasma CFH concentration as the quantitative trait and included other potential covariates including sex, age, age-by-sex, smoking and BMI. The polygenic model incorporates the family structure and covariates to estimate the heritability of the trait and the proportion of trait variance explained by the covariates.

Addition of one copy of the minor allele, T (His^{402}) of the *CFH* Y402H variant resulted in a decrease of 18.9 μg/ml (± 2.78, SE) in the mean plasma CFH concentration (P = 1.2E-11) (**Table 6.2**). Sex, age-by-sex interaction and BMI also showed significant association with plasma CFH concentrations.

Male sex was associated with a decrease of 11.2 μg/ml (± 3.68, SE) in the mean plasma CFH concentration (P = 0.002). Age was not a significant covariate when analysed on its own (P = 0.193) (**Table 6.2**). A significant association was also observed between plasma CFH and BMI. Each unit increase in BMI was found to be associated with a 3.3 μg/ml (± 0.44, SE) increase in mean plasma CFH (P = 1.1E-12) (**Table 6.2**). A significant interaction was observed between age and sex, with each one year increase in age of males being associated with a 0.90 μg/ml (± 0.95, SE) mean increase in mean plasma CFH concentration (P = 0.003). Smoking did not show any significant association with changes in plasma CFH (P = 0.642).

Plasma CFH concentration also showed a significant narrow-sense heritability, estimated in the Vis population as 0.46 ± 0.13 (P = 0.00058). The proportion of the phenotypic variance accounted for by significant covariates was estimated to be 0.204. This confirmed the presence of other genetic and/or environmental factors which influence plasma CFH.

6.4.2 Plasma CFH distribution in the isolated Rucphen (Dutch) community

CFH measurements were carried out in 490 plasma samples of unrelated individuals from an isolated Dutch population, using the CFH ELISA described previously. Information on the age and sex of the individuals screened were reported in **Table 2.1**, which was used for the analysis of covariates.

Measurements of plasma CFH were normalised using internal controls as described before. The measured plasma CFH showed a two-fold range of concentration (**Figure 6.8-B**). The mean plasma CFH concentration in the Rucphen population was 509.7 μg/ml (± 56.3, SD). CFH measurements in this population also showed a normal distribution with a kurtosis value of 0.11 ± 0.22.

6.4.2.1 Analysis of covariates

Tests for association between covariates and plasma CFH concentration were carried out in the Dutch Rucphen samples by univariate analysis using a general linear model implemented in the statistical software SPSS v12. Age, sex, age-by-sex interaction and SNP rs1061170 (*CFH* Y402H) genotypes (**Table 6.4**) were tested as possible covariates in the analysis.

Both rs1061170 and age had significant effects on plasma CFH concentration in the Rucphen population (**Figure 6.9**). Mean plasma CFH decreased by 13 μg/ml for each copy of SNP rs1061170 minor allele using an additive model (P = 4.8E-04). Age was also found to be a significant covariate since plasma CFH concentration increased with age (β = 0.57 μg/ml per one year increase in age, P = 2.9E-03). Sex was not found to be a significant covariate for plasma CFH concentration in this population. The proportion of the phenotypic variance explained by the two significant covariates, age and SNP rs1061170, was estimated at 0.039.

6.4.3 Analysis of plasma CFH in the Scottish AMD series

Plasma CFH measurements were carried out in 351 Scottish AMD cases (mean age = 75.6 years, 62.8% females) and 198 age and sex-matched controls (mean age = 73.1 years, 62.2% females) (**Table 2.1**, Chapter 2).

6.4.3.1 Analysis of plasma CFH covariates in the Scottish AMD cases

The distribution of plasma CFH in cases of AMD was found to be leptokurtic and showed a 4.2-fold range of values (**Figure 6.8-C**). The mean unadjusted plasma CFH concentration in the AMD cases was 412.6 μg/ml.

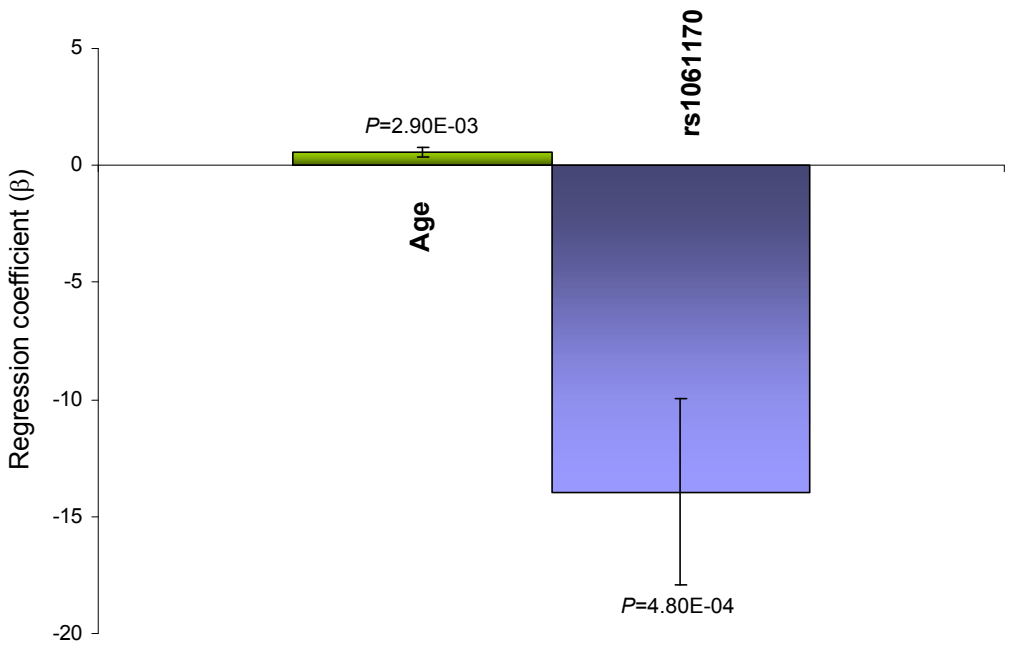


Figure 6.9 Plasma CFH covariate analysis in the isolate Dutch population.

Diagram showing significant changes in mean plasma CFH associated with significant covariates in a linear regression analysis. The y-axis shows changes in the mean plasma CFH associated with age and SNP rs1061170 (*CFH* Y402H). Error bars are standard error of the mean. *P*-values are shown for each covariate effect.

SNP	MAF	Genotypic Frequency (%)	HWD *P*-value
rs1061170	0.32	TT (47.0), TC (41.6), CC (11.4)	0.28

Table 6.4 Descriptive statistics for SNP rs1061170 in the Dutch Rucphen population.

SNP rs1061170 (*CFH* Y402H) allele and genotypic frequencies measured in 500 unrelated individuals from the isolated Dutch Rucphen population. Deviation from Hardy Weinberg equilibrium was also assessed.

MAF: minor allele frequency, HWD *P*-value was calculated for deviation from Hardy-Weinberg equilibrium.

Available covariates including age, sex, age-by-sex, smoking and *CFH* rs1061170 (Y402H) were tested for association with plasma CFH concentration using the general linear model implemented in the software SPSS v.12. None of the potential covariates used were found to be statistically significant.

6.4.3.2 Analysis of plasma CFH covariates in the Scottish AMD control sample

The control population had a platykurtic distribution with a 1.8-fold range (**Figure 6.8-C**). The mean unadjusted plasma CFH concentration in the Scottish control series was 428.3 μg/ml.

Association tests were carried out for potential covariates including sex, age, age-by-sex, smoking and SNP rs1061170 (*CFH* Y402H) using the general linear model implemented in the statistical software package SPSS v.12. The results indicated that age was the only significant covariate for plasma CFH with a reduction of 18.2 μg/ml in the mean CFH concentration associated with each year of life (P = 0.031) (**Figure 6.10**). No significant associations were observed with other covariates in the model.

6.4.3.3 Comparison of plasma CFH concentrations between Scottish AMD cases and controls

Plasma CFH concentrations were compared between 351 cases of AMD and 198 disease-free controls using the statistical software package SPSS v.12.

Mean unadjusted plasma CFH in the AMD cases was 412.6 μg/ml, 15.7 μg/ml lower than the Scottish control population (**Figure 6.8-A** and **Figure 6.11**). An independent-samples two-tailed t-test was used to compare the AMD cases and controls and the results showed that there were significant differences in plasma CFH concentration between AMD cases and controls after adjusting for age ($P_{\text{two-tailed}}$ = 0.003) (**Figure 6.11**). The proportion of the variance in plasma CFH concentration explained by disease status, represented as R^2, was found to be 0.016. A binary logistic regression analysis of plasma CFH concentration divided into quartiles, showed an AMD odds ratio of 2.1 (95% C.I. 1.3 - 3.4, P = 0.003) associated with the lowest quartile (0 - 25%) compared with the highest quartile (75% - 100%).

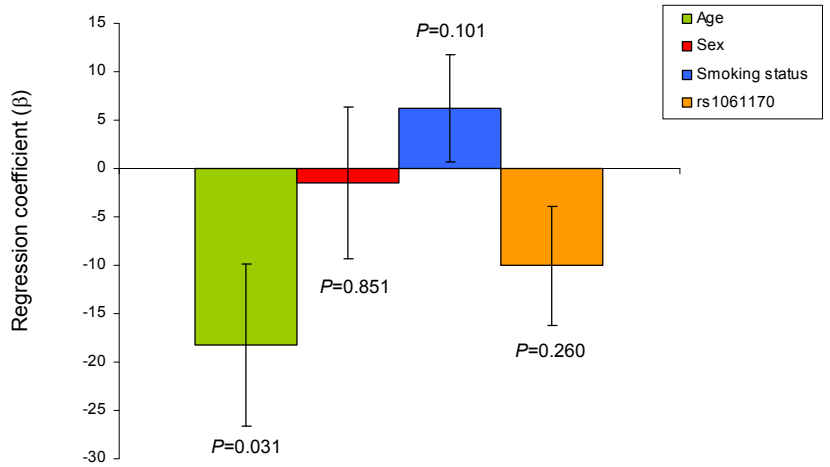
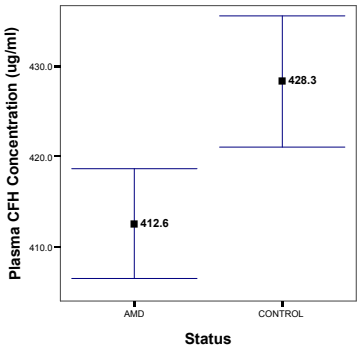


Figure 6.10 Analysis of plasma CFH covariates in an urban Scottish population.

Diagram showing the association between plasma CFH concentration and a number of covariates in 187 control individuals of Scottish origin. Age was found to be the only significant covariate which was associated with a decrease in mean plasma CFH as shown on the y-axis. Error bars are standard errors of the mean.



Covariates	Type III Sum of Squares	d.f.	F	*p*	R^2
Sex	0.126	1	0.081	0.775	0.000
Age	1.157	1	0.750	0.387	0.001
Disease status	13.870	1	8.989	0.003	0.016

Figure 6.11 Analysis of plasma CFH concentrations between Scottish AMD cases and controls.

Diagrammatic presentation of the plasma CFH differences between cases and controls in the Scottish AMD cohort (top). Table shows the disease status used as a covariate for plasma CFH levels in a general linear model. Error bars represent 95% confidence interval of the mean. d.f.: degrees of freedom. F: F test statistic with the *P*-value for the test. R^2 shows the proportion of the variance in plasma CFH concentration explained by the covariates.

6.4.4 Analysis of plasma CFH in a Japanese AMD series

Quantitation of plasma CFH was carried out in 30 AMD cases (mean age of 77.3 years, 16.7% female) and 58 age-matched controls (mean age of 76.6 years, 60.3% female) of Japanese origin. Mean plasma CFH was found to be 326.8 μg/ml (± 61.50, SD) in cases and 337.6 μg/ml (± 68.12, SD) in controls (**Figure 6.8-D**).

Age, sex, age-by-sex and *CFH* rs1061170 (Y402H) were used as covariates in a general linear model to assess their effects on changes in the mean plasma CFH concentration in AMD cases and controls separately, using the statistical software SPSS v.12. Both case and control distributions were checked for normality, using the Kolmogorov-Smirnov test and both were consistent with a normal distribution ($P > 0.01$). None of the potential covariates were found to be associated with plasma CFH in the Japanese cases or controls ($P > 0.05$).

The minor allele frequency of the *CFH* rs1061170 (Y402H) variant was 13.3% and 4.3% in the AMD cases and controls, respectively. The difference was found to be significant ($P = 0.037$).

An independent-sample two-tailed t-test was then used to compare the means of plasma CFH concentrations in AMD cases and controls, using the software SPSS v.12. The assumption of homogeneity of variances of the two groups was tested using Levene's test which did not reject the hypothesis that the variances are equal in cases and controls (F = 0.136, $P = 0.713$) (**Table 6.5)**. The t-statistic was then calculated for the combined groups, as the ratio of the difference between sample means divided by the standard error of the difference. The results showed no significant difference between the means of plasma CFH between Japanese AMD cases and controls ($P_{\text{two-tailed}} = 0.469$) (**Table 6.5**).

6.4.5 Comparison of plasma CFH in general populations of Vis, Rucphen, Scotland and Japan

Plasma CFH concentrations were compared between samples of Croatian (N = 1,004), Dutch (N = 490), Scottish (disease-free controls) (N = 198) and Japanese (AMD-free controls) (N = 58) general populations. Comparisons were carried out after adjusting for the effect of the significant covariates identified earlier.

	Levene's Test for Equality of Variances		t-test for Equality of Means				
	F	***P***	**T**	**d.f.**	$P_{\text{two-tailed}}$	**Mean Difference**	**Std. Error Difference**
Plasma CFH (μg/ml)	0.136	0.713	-0.728	86	0.469	-10.79	14.834

Table 6.5 Analysis of plasma CFH concentration in the Japanese AMD cohort.

Mean plasma CFH concentrations were compared between 30 AMD cases and 58 controls from a Japanese population. An independent-sample two-tailed t-test was used for the analysis after assessing the homogeneity of the group variances using Levene's F-test. Two-tailed t-test and *P*-value for the comparison of mean CFH concentrations in AMD cases and controls were calculated with 86 degrees of freedom (d.f.). Analysis was carried out using statistical software package SPSS v.12.

The unadjusted mean CFH concentration in the Vis population was 109.5 μg/ml lower than the mean value obtained in the Rucphen (Dutch) population (P = 1.9E-178). After adjusting for covariate effects associated with age, sex and *CFH* rs1061170, the difference in the mean CFH concentrations between the Vis and Rucphen populations was no longer significant (P = 1.00).

The unadjusted mean CFH concentration in the Scottish control population was 81.4 μg/ml lower than the mean plasma CFH in the Rucphen (Dutch) population (P = 3.4E-57). After adjusting for covariates age and *CFH* rs1061170, this difference was also found to be non-significant (P = 1.00).

The unadjusted mean CFH concentration in the Scottish control population was 28.1 μg/ml higher than the unadjusted mean in the Vis population (P = 7.9E-11). However, after adjusting for covariate effects associated with age, sex and *CFH* rs1061170, this difference was also found to be non-significant after adjustments (P = 1.00).

Finally, the difference in the mean plasma CFH concentration between the general population of Vis and the Japanese control population was found to be non-significant (P = 1.00). The difference in the mean plasma CFH between the Scottish and the Japanese control populations was also found to be non-significant after adjusting for the effects associated with age and *CFH* rs1061170 (P = 1.00).

6.5 Genome-wide linkage analysis of plasma CFH

Genome-wide analysis of linkage was carried out using the variance components methodology implemented in the software SOLAR (Almasy & Blangero, 1998). The genome-wide linkage scan consisted of 810 STR markers genotyped in 1,042 individuals from the island of Vis. The series consisted of randomly selected individuals, 58% of whom were members of 122 pedigrees. A full description of the samples was presented in **Table 2.1**. Microsatellite markers were spaced approximately 5 cM apart. A full description of the linkage scan and the underlying principles were described in Section 2.5.8.

The null hypothesis being tested was that the additive genetic variance due to a putative quantitative trait locus (QTL) equalled zero and so only the polygenic

effects (due to pedigree relationship) on the phenotype were modelled. The hypothesis of no linkage was tested by comparing the likelihood of the null polygenic model, including the significant covariates of sex, age-by-sex interaction, BMI and SNP rs1061170 (*CFH* Y402H), with that of a model that estimated the additive genetic variance at the putative QTL.

The results of the linkage scan for all 22 autosomal chromosomes are summarised in **Figure 6.12**. The highest LOD score detected was 1.16 on chromosome 10. No multipoint LOD scores indicative ($Z \geq 3$) or suggestive ($Z \geq 2$) of linkage were observed in this analysis.

6.6 Summary and discussion

A high throughput "sandwich" ELISA was developed and optimised for the quantitation of human complement factor H in plasma. The possibility of the polyclonal capture antibody detecting the CFH-related proteins in plasma (as a result of competition with full length CFH) cannot be excluded despite the western blot analysis of the detection antibody showing high specificity to the full length CFH. Determining the proportion of the quantitated protein which is accounted for by the related proteins will therefore require further investigation.

Inter and intra-experimental variation was assessed and the coefficients of variation were all found to be below 5%. Analysis of assay results and variation after one or two sample freeze-thawing demonstrated the ability of plasma CFH to withstand freeze-thawing and the stability of the protein at 4^{o}C for up to four days.

Analysis of plasma CFH in the Vis population

Plasma CFH measurements were carried out in four populations. The concentrations detected in 1,004 individuals from the Croatian isolate of Vis showed a three-fold range in values (218 - 654 μg/ml). The range of values differed slightly from what has previously been reported in an urban Spanish population (100 - 600 μg/ml) (Esparza-Gordillo *et al*., 2004). This could be explained by reduced genetic and environmental variation in this isolate population as a result of geographical isolation (Rudan, 1999a; Polasek *et al*., 2006).

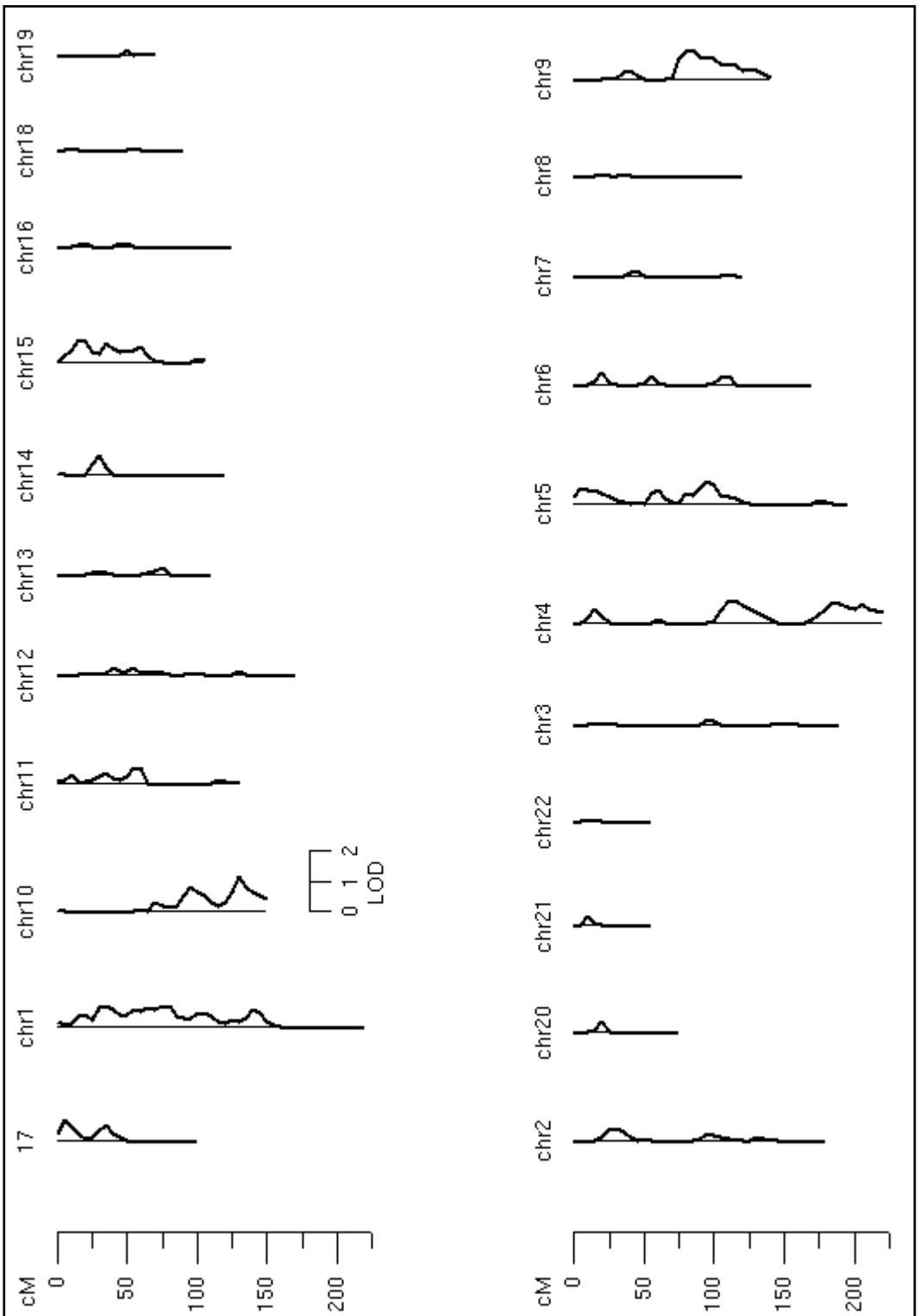


Figure 6.12 Results of plasma CFH linkage analysis in the Vis population.

LOD scores were obtained using a genome-wide linkage analysis of 810 microsatellite markers in 1,042 members of the Vis population. Markers were spaced at approximately 5 cM intervals. The model for linkage was based on a polygenic model consisting of BMI, SNP rs1061170, sex and age-by-sex interaction which were found to be significant covariates for plasma CFH. Analysis was carried out using the software package SOLAR (Almasy& Blangero, 1998).

The demographic history of the island of Vis dates back to approximately 1,000 years BC. Since then the Vis population has gone through major admixture and bottleneck events; the former occurred due to immigration from the Croatian mainland (17th century) and the latter as a result of emigration (44 - 88% reduction in population size) in the 20th century, hence lowering the genetic heterogeneity (Vitart *et al*., 2006). Variation in plasma CFH measurements between the Vis population and other published populations can also be explained by assay differences, including the purity of the CFH standard and the efficiency of the capture and detection antibodies. The purity of the standard was assessed by silver staining and the specificity of the antibody was tested using a western blot (**Figure 6.2**).

Test of association between plasma CFH concentration and a number of covariates were assessed using a general linear regression model in SPSS v.12. The *CFH* Y402H (rs1061170) risk allele (His^{402}) which has been shown to increase susceptibility to AMD (Haines *et al*., 2005; Edwards *et al*., 2005; Klein *et al.*, 2005, Zareparsi *et al*., 2005a) was associated with significantly lower plasma CFH concentrations in the Vis general population (P = 1.2E-11, **Table 6.2**). This amino acid substitution could affect levels of circulating CFH by disrupting the stability, solubility or secretion of the protein from its main site of production in the liver.

BMI was associated with a small increase in mean plasma CFH (3.3 μg/ml ± 0.44, SE) which was highly significant (P = 1.1E-12) and explained 8.7% of the unadjusted phenotypic variance (**Table 6.2**). Raised BMI (≥ 30) has been shown to be associated with increased risk of AMD, independently of *CFH* variant Y402H, with an odds ratio (OR) of 2.1 (Seddon *et al*., 2006a). Since low rather than high circulating CFH concentrations are expected to increase the risk of AMD, this contradictory finding is difficult to explain unless there is some confounding factor, such as age, that influences plasma CFH. However, age alone was not associated with plasma CFH. The concentration of CFH in plasma has been shown to increase with age (Esparza-Gordillo *et al*., 2004). One explanation for the lack of association between plasma CFH and age in the Vis population is that the effect of age on plasma CFH is accounted for by BMI. This was tested and confirmed by removing BMI from the univariate analysis of plasma CFH which resulted in statistically significant association between age and plasma CFH concentration (P < 0.001). Sex

and an age-by-sex interaction were found to be significantly correlated with plasma CFH concentrations in the Vis population (P_{sex} = 0.002, $P_{age*sex}$ = 0.003). Age and sex have been shown to significantly correlate with incidence of macular degeneration and the disease is more common amongst older women (Maltzman *et al*., 1979). According to Evans (2001) age and genetic susceptibility are the two most important risk factors in AMD (Evans, 2001).

Smoking was not significantly correlated with changes in plasma CFH concentration in the Vis population ($P > 0.05$) (**Table 6.2**). However, smoking has been shown to be associated with increased risk of AMD in older individuals (Seddon *et al*., 2006a). Furthermore, smoking has a well-established role in the stimulation of pro-inflammatory responses and is associated with increased levels of inflammatory markers (Bakhru & Erlinger, 2005).

The heritability of plasma CFH concentration in the Croatian isolate population of Vis was estimated to be 0.46. However, this heritability estimate is limited to additive gene effects and excludes non-additive factors such as dominance and epistasis, which makes this a conservative estimate of overall heritability. Sex, age-by-sex interaction, BMI and the AMD susceptibility SNP rs1061170 (Y402H) together explained 20.4% of the plasma CFH variance in this population. This confirms the presence of other genetic or environmental factors which influence plasma CFH and hence potentially affect susceptibility to AMD.

Analysis of plasma CFH in the Rucphen (Dutch) population

Plasma CFH concentration was measured in 490 unrelated individuals from the Rucphen (Dutch) community used as a replication series. The results showed a two-fold range of concentration and a mean of 509.7μg/ml (**Figure 6.8-B**).

Covariate analysis included age, sex, an age-by-sex interaction and the *CFH* rs1061170 (Y402H). Age was found to be a significant covariate for plasma CFH concentration in this population (P = 2.9E-03) and an increase in age resulted in increased plasma CFH concentration (β = 0.57 μg/ml per year). BMI data were not available for these samples. Similar to the results of the Vis population, the *CFH* Y402H (rs1061170) risk allele (His^{402}) was associated with significantly lower plasma CFH concentrations in the Rucphen population (P = 4.8E-04) (**Figure 6.9**).

Analysis of plasma CFH in the Scottish AMD series

Analysis of plasma CFH was carried out in 351 unrelated Scottish cases of AMD and 198 age- and sex-matched AMD-free controls.

A number of covariates including age, sex, age-by-sex interaction, smoking and *CFH* rs1061170 (Y402H) were analysed for potential associations with plasma CFH concentrations. In the Scottish cases of AMD, none of the potential covariates were found to be significantly associated with plasma CFH. In the Scottish control population, age was the only significant covariate for plasma CFH, associated with a reduction of 18.2 μg/ml in the unadjusted mean CFH concentration per year increase in age ($P = 0.031$) (**Figure 6.10**). BMI data were not available for the Scottish AMD case-control samples. Lack of association between plasma CFH concentrations and sex, smoking and *CFH* rs1061170 (Y402H) in the Scottish control population could be explained by ascertainment bias as the study sample had already been selected based on the absence of disease.

Comparison of plasma CFH concentrations between Scottish AMD cases and controls showed that CFH concentrations were significantly higher in the control population compared with cases of AMD ($P = 0.003$) after adjusting for age (**Figure 6.11**). Furthermore, binary logistic regression analysis of plasma CFH concentration divided into quartiles, showed an AMD odds ratio of 2.1 (95% C.I. 1.3 - 3.4, $P = 0.003$) associated with the lowest quartile (0 - 25%) compared with the highest quartile (75% - 100%). AMD disease status was found to significantly explain 1.6% of the variation in plasma CFH concentration ($P = 0.003$). Identification of plasma CFH concentration as a risk factor for AMD is a novel finding. Circulating CFH concentration appears to be one of many factors which influence susceptibility to age-related macular degeneration. Whether the difference between cases and controls is a cause or effect requires further investigation however, low plasma CFH is plausibly associated with increased activation of the alternative complement pathway and increased susceptibility to AMD.

Analysis of plasma CFH in the Japanese AMD series

Plasma CFH measurements were carried out in 30 cases of AMD and 58 age-matched controls. Mean plasma CFH was found to be 326.8 μg/ml in the cases of AMD and 337.6 μg/ml in the controls (**Figure 6.8-D**).

Plasma CFH concentrations did not differ significantly between AMD cases and controls in the Japanese population (P = 0.469) (**Table 6.5**). The small number of samples entails a lack of power to detect an association between predictors and plasma CFH concentration and to find significant differences between AMD cases and controls in this population.

It is interesting that despite a small sample size, the allele frequency differences for *CFH* SNP rs1061170 were significant between the Japanese cases and controls despite the lack of association between this SNP and AMD reported previously (Uka *et al*., 2006; Gotoh *et al*., 2006; Okamoto *et al*., 2006). There may be a reduced frequency of polymorphisms other than rs1061170 (Y402H) which affect AMD by influencing plasma CFH concentrations in the Japanese population.

Comparison of the results between the Croatian, Dutch, Scottish and Japanese populations

Reduced genetic heterogeneity as a result of isolation (Rudan *et al*., 1987; Forenbaher *et al*., 1999), genetic drift, founder effects (Smoljanović *et al*., 2006; Rudan *et al*., 2006; Jeran & Havas, 2007) or selection at the *CFH* locus due to previous history of infections (Hageman *et al*., 2006) may have resulted in the higher frequency of protective alleles (associated with higher plasma CFH concentration) in the isolate populations of Vis and Rucphen. The unadjusted mean CFH concentration in the isolated Dutch Rucphen population was approximately 100 μg/ml higher than that obtained in the Croatian samples which was found to be statistically significant (P = 1.9E-178). However, after adjusting for the covariate effects associated with age, sex and *CFH* rs1061170, this difference was found to be non-significant (P = 1.00). The mean age was found to be significantly different between the Vis (56.1 ± 0.49, SE) and the Rucphen (59.2 ± 0.62, SE) populations. Furthermore, the minor allele frequency of *CFH* rs1061170 (Y402H) was found to be significantly different between the Vis (MAF = 0.39) and the Rucphen (MAF =

0.32) populations using a Fisher's exact test (P = 1.5E-03). This implies that the differences observed between plasma CFH measurements in the two populations are due to differences in age distributions and the minor allele frequency of the *CFH* rs1061170 variant which was found to be significantly associated with plasma CFH in both populations.

The mean plasma CFH concentration in the Scottish control population was 28.1 μg/ml and 81.4 μg/ml lower than the values obtained in the Vis (P = 7.9E-11) and Rucphen (P = 3.4E-57) populations. However when age, sex and *CFH* rs1061170 were included as covariates, the difference in plasma CFH concentrations between the two populations was no longer significant (P = 1.00). While no statistically significant differences were observed in the minor allele frequency of the *CFH* rs1061170 (Y402H) polymorphism between the Vis population (MAF = 0.39) and the Scottish urban population (MAF = 0.34) ($P > 0.05$), age distributions were found to significantly differ between the two populations (P = 3.5E-29). Therefore, this could have resulted in differences in plasma CFH distributions.

Statistically significant differences in the unadjusted mean plasma CFH concentrations were observed between the Japanese control series and the Vis population (P = 3.1E-13), the Dutch population (P = 1.1E-74) and the Scottish AMD-free control population (P = 7.9E-23).

Genome-wide linkage analysis of plasma CFH in Vis

It has been suggested that the probability of obtaining a QTL through linkage depends on the heritability of the trait attributable to that QTL (Almasy & Blangero, 1998). Based on this hypothesis, a genome-wide linkage screen of plasma CFH using 810 microsatellite markers was performed in 1,042 individuals from Vis in order to identify genomic loci which regulate plasma CFH concentration. No LOD scores indicative (LOD ≥ 3) or suggestive (LOD ≥ 2) of linkage were observed in the linkage analysis (**Figure 6.12**). The power to detect a QTL by linkage depends on the proportion of variation accounted for by the QTL, the number of families and the family size and structure (Atwood & Heard-Costa, 2003). The small average QTL effect size underlying the variation in plasma CFH is probably the main reason for the lack of significant evidence of linkage. Predictive analysis of power to detect

QTL by variance components methodologies has been assessed based on the family structure in the Vis data set and the results have shown that sufficient power is only available for the detection of QTL of at least 30% effect size (Dr. V. Vitart, personal communication).

In conclusion, analysis of plasma CFH concentration using a novel CFH ELISA identified significant associations between the *CFH* SNP rs1061170 (Y402H) and plasma CFH in the Vis and Rucphen data sets (**Table 6.2** & **Figure 6.9**) suggesting that there may also be other variants within the *CFH* locus which influence plasma CFH concentration. The SNP rs1061170 could be affecting the concentration of CFH in plasma by altering the stability of the protein or its secretion from the liver. A second novel finding is the identification of plasma CFH as a risk factor for AMD since significant differences in plasma CFH concentrations were observed between Scottish AMD cases and controls (**Figure 6.11**). Furthermore, a narrow-sense heritability of 0.46 was estimated for plasma CFH in the Vis population suggestive of underlying additive genetic variants influencing this trait. Multipoint linkage analysis of plasma CFH did not identify suggestive or significant linkage peaks, due to low power in linkage (**Figure 6.12**). However, as the predictive analysis of power showed sufficient power is available in the Vis data to detect variants of effect size as low as 3% using association mapping. Therefore, a high-resolution genome-wide association analysis was carried out to identify genomic loci influencing plasma CFH.

CHAPTER 7

GENETIC ASSOCIATION ANALYSIS OF PLASMA CFH

7.1 Introduction

In this chapter, the genetic analysis of the *CFH* locus and results of a genome-wide association analysis are presented. Following identification of the *CFH* gene as a major AMD genetic susceptibility locus (Haines *et al*., 2005; Edwards *et al*., 2005; Klein *et al.*, 2005) a number of studies identified other genetic susceptibility factors in the region. Haplotypes consisting of multiple SNPs excluding the *CFH* Y402H variant were shown to have a stronger association with disease susceptibility than Y402H-containing haplotypes (Li *et al*., 2006). Furthermore, a common haplotype carrying a deletion of approximately 86 kb containing the *CFHR3* and *CFHR1* genes was found to be associated with decreased risk of AMD, acting independently of Y402H (Hughes *et al*., 2006).

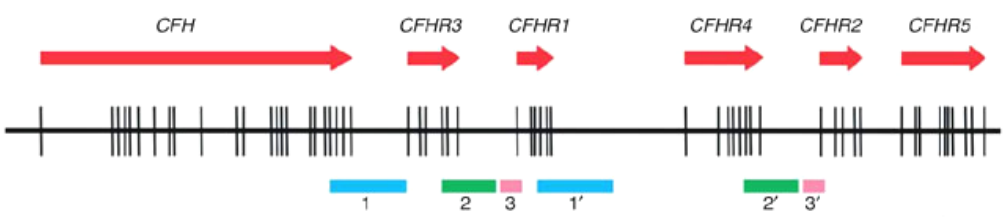


Figure 7.1 Diagrammatic representation of the *CFH* locus.

Diagram showing the *CFH* locus on chromosome 1q32 spanning approximately 357 kb of DNA. Five *CFH*-related genes are also shown (*CFHR1-5*). Blue, green and pink blocks below the black line represent three areas of genomic duplication with more than 96% sequence homology. Red arrows show the direction of transcription and exons are shown as black vertical lines (Figure adapted from Hughes *et al.*, 2006).

Following the identification of plasma CFH as a biomarker for AMD, and the association between the *CFH* SNP rs1061170 (Y402H) with plasma CFH, a genome-wide association analysis using 317,503 SNPs (Illumina HumanHap300 Array®) was carried out with the aim of identifying polymorphisms which affect plasma CFH concentration, hence influence susceptibility to AMD. Polymorphisms identified in previous published work to influence susceptibility to AMD and not included in the Illumina Hap300 array were genotyped by TaqMan® assays. The results of

genotyping the *CFHR3/CFHR1* deletion were also included in the association analysis.

7.2 Results

7.2.1 Genotyping of the *CFHR3/CFHR1* deletion

A common deletion polymorphism of *CFHR3/CFHR1* (**Figure 7.1**) has been found to significantly protect against AMD independently of the *CFH* Y402H variant (Hughes *et al*., 2006). The deletion is present on a protective haplotype which is present in 7.8% of AMD chromosomes and 20% of chromosomes of control individuals. The working hypothesis in the present study was that the deletion polymorphism confers a protective effect against AMD through its association with other variants directly increasing plasma CFH concentration.

A high throughput real-time quantitative PCR (qPCR) was developed to genotype the *CFHR3/CFHR1* deletion polymorphism as described in Methods.

7.2.1.1 Results of *CFHR3/CFHR1* deletion genotyping

Genotyping of the *CFHR3/CFHR1* deletion was carried by calculating the difference in threshold cycle (C_t) values observed for amplification of the *CFHR3* (FAM fluorescence) and *HBB* (VIC fluorescence) genes (**Figure 7.2**). The observed C_t value is proportional to the number of target genes. Theoretically, a unit difference in C_t would indicate a loss or gain of one copy number of the template. A difference in C_t value of exactly one unit between *CFHR3* and *HBB* was rarely obtained. This is believed to be the result of varying efficiencies of primers and probes in annealing to the DNA template. Therefore, a C_t difference of zero to 0.4 was taken as a homozygous non-deleted, and a difference greater than 0.7 was taken as a heterozygous genotype. Homozygous deleted genotypes resulted in no amplification of the *CFHR3* template. The genotypes of individuals homozygous for the *CFHR3/CFHR1* deletion were confirmed in a duplex PCR in which the absence of a 219 bp amplicon of *CFHR3* in the presence of a 280 bp amplicon of *HBB* represented a homozygous deleted genotype (**Figure 7.3**). Heterozygous non-deleted or homozygous non-deleted genotypes were shown by the presence of the 219 bp amplicon of *CFHR3* (**Figure 7.3**).

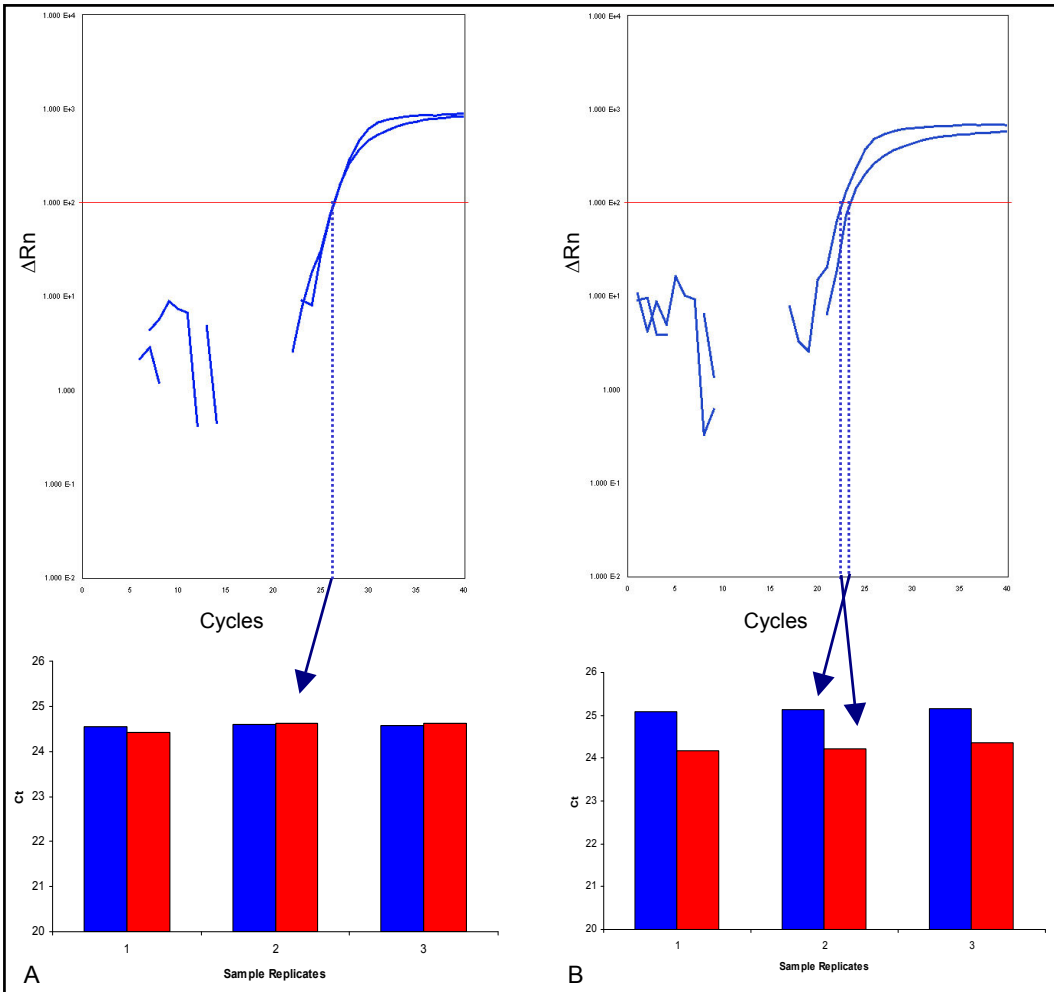


Figure 7.2 Genotyping of the *CFHR3/CFHR1* deletion by quantitative real-time PCR.

Examples of a homozygous non-deleted **(A)** and a heterozygous **(B)** genotype are shown. Amplification plots captured in real-time are shown on the top. Extracted C_t values used to deduce each genotype are presented at the bottom (blue: FAM-labelled *CFHR3* probe, red: VIC-labelled *HBB* probe) . The C_t value is where the vertical dotted line crosses the x-axis. Each reaction was carried out in triplicate.

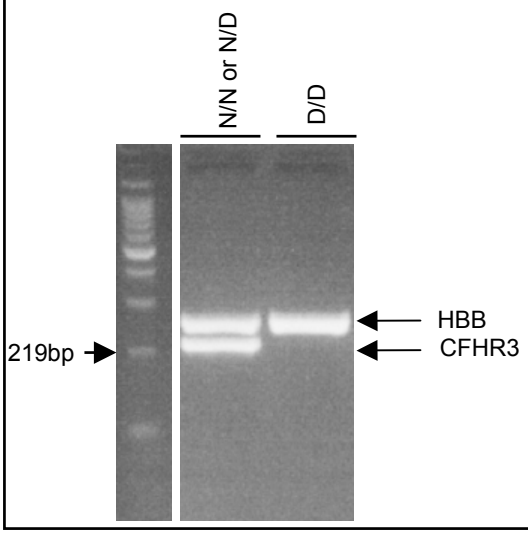


Figure 7.3 Screening for individuals homozygous for the *CFHR3/CFHR1* deletion.

A sample that is homozygous for the deletion with a single band at *HBB* (D/D)and a sample with ambiguous genotype with a band present for both *HBB* and *CFHR3* (N/N or N/D). Each well contains 10 μl of PCR product run on a 2% agarose mini-gel. A 100 bp DNA marker is shown on the left. D/D, homozygous for deletion; N/N, homozygous non-deleted; N/D, heterozygous for deletion.

Genotyping of the *CFHR3/CFHR1* deletion was carried out in triplicate using DNA from individuals from the Croatian island of Vis, from the Dutch Rucphen isolate and from the Scottish general population series of AMD cases and controls (**Table 7.1**). Genotypes of 472 individuals in pedigrees from Vis were checked for Mendelian inheritance errors using the software package PLINK (Purcell *et al*., 2007) and none were found. Allele frequencies were consistent with Hardy-Weinberg equilibrium using a threshold of $P_{HWD} < 1.0E-04$. The frequency of the *CFHR3/CFHR1* deleted allele did not differ significantly between the Scottish non-AMD controls (MAF = 0.206) and the Croatian isolate population (MAF = 0.197) ($P = 0.592$) (**Table 7.1**). The minor allele frequency was significantly higher in the Dutch Rucphen isolate (MAF = 0.295) compared with Vis ($P = 1.1E-06$). The *CFHR3/CFHR1* deletion minor allele frequency observed in Scottish cases of AMD (MAF = 0.132) was 36% lower than in the Scottish controls ($P = 1.4E-03$) (**Table 7.1**). The frequency of the deleted allele in the AMD cases was 58% lower than that observed in Croatian sample set ($P = 7.2E-04$) (**Table 7.1**).

7.2.2 QTL Association analysis of CFH in Vis

Genome-wide association analysis of plasma CFH was carried out using the genome-wide association tool set PLINK (Purcell *et al*., 2007). Test of association was performed using 317,509 polymorphisms consisting of 317,503 SNPs from the Illumina Hap300 array and 6 bi-allelic markers genotyped independently, as described before. The aim was to identify quantitative trait loci which affect plasma CFH concentration and to examine their influence on susceptibility to AMD.

7.2.2.1 *CFH* SNP genotyping

Six polymorphisms, including the *CFHR3/CFHR1* deletion were genotyped using either the TaqMan® allelic discrimination assay discussed above or by direct sequencing (**Table 7.2**). The aim was to assess the association between plasma CFH concentration and SNPs reported by Li and colleagues (2006) to influence susceptibility to AMD.

		Croatian (N=932)	Dutch (N=866)	Scottish AMD Cases (N=726)	Scottish AMD Controls (N=384)
CFHR3/CFHR1 Deletion Frequency (%)	D/D	3.9	5.8	2.5	4.7
	N/D	33.3	47.6	21.5	31.8
	N/N	62.8	46.6	76.0	63.5
	MAF	19.7	29.5	13.2	20.6

Table 7.1 Genotype and allele frequencies for the deletion of the *CFHR3/CFHR1* in four population sample sets.

Summary of the genotype and allele frequencies of the *CFHR3/CFHR1* deletion using unrelated individuals from the Croatian and Dutch isolates, and a cohort of Scottish AMD cases and disease-free controls. *N* shows the number of chromosomes for which the genotyping was carried out.

SNP ID	Minor Allele	MAF	*P* (HWD)	*N*
rs2274700	T	0.421	0.777	948
rs1048663	A	0.155	0.861	956
rs412852	C	0.418	1.000	954
rs1066420	G	0.145	1.6E-04	958

Table 7.2 Details of additional (non-Illumina) *CFH* SNPs genotyped in the Vis (Croatian) sample set.

Summary of allele frequencies observed for the *CFH* SNPs. These SNPs represent the high risk haplotype described by Li and colleagues (Li *et al*., 2006; Table 2), except for the SNP rs1066420 which has replaced rs128054 in the database (NCBI Build 35 assembly) and was therefore used in the present study. *P*-values detected for deviations from Hardy Weinberg equilibrium are shown. MAF, minor allele frequency; HWD, Hardy-Weinberg disequilibrium; *N*, number of chromosomes.

A haplotype was identified by Li and colleagues (2006) to affect susceptibility to AMD independently of *CFH* Y402H. This haplotype consisted of five SNPs, rs2274700, rs1048663, rs412852, rs1280514 and rs11582939. Of these, *CFH* rs2274700 showed stronger genome-wide association with AMD than *CFH* Y402H (Li *et al*., 2006). Following the publication of the report by Li *et al*., the SNP rs1280514 was replaced by rs1066420 in the database (NCBI Build 35 assembly). Therefore, SNP rs1066420 was used in the present study.

We aimed to identify the effect on plasma CFH concentration associated with *CFH* rs2274700 and the 5-SNP haplotype. SNPs rs2274700, rs1048663, rs412852 and rs1066420 were genotyped as described before (**Table 7.2**). SNP rs11582939 was included in the Illumina Hap300 array. All genotypes were checked for Mendelian inheritance errors and none were found. Deviations from Hardy-Weinberg equilibrium were also excluded. *CFH* rs1066420 was the only SNP which showed significant deviation from HWE (P_{HWD} = 1.6E-04). Direct sequencing of samples with ambiguous genotypes was used to exclude the possibility of genotyping errors, and this marker was consequently removed from further analysis. The only difference in *CFH* SNP allele frequencies observed between those genotyped in Vis and those reported in HapMap was rs2274700, which had a MAF of 0.42 compared with 0.50 in HapMap.

7.2.2.2 Genome-wide QTL association analysis of CFH

Genome-wide association analysis of 317,508 markers was carried out using 966 individuals from the island of Vis, 58% of whom were members of 122 families (**Figure 7.4**). Twenty-five individuals with missing phenotypes were removed, leaving 941 in the final analysis. 8,984 SNPs were excluded because of low genotyping call rates (<90%) and 379 SNPs were removed due to minor allele frequencies <1% leaving 308,144 SNPs. The genome-wide significance threshold was set at $-\log_{10}(P) = 6.8$ after correction for 308,144 tests using the Bonferroni correction.

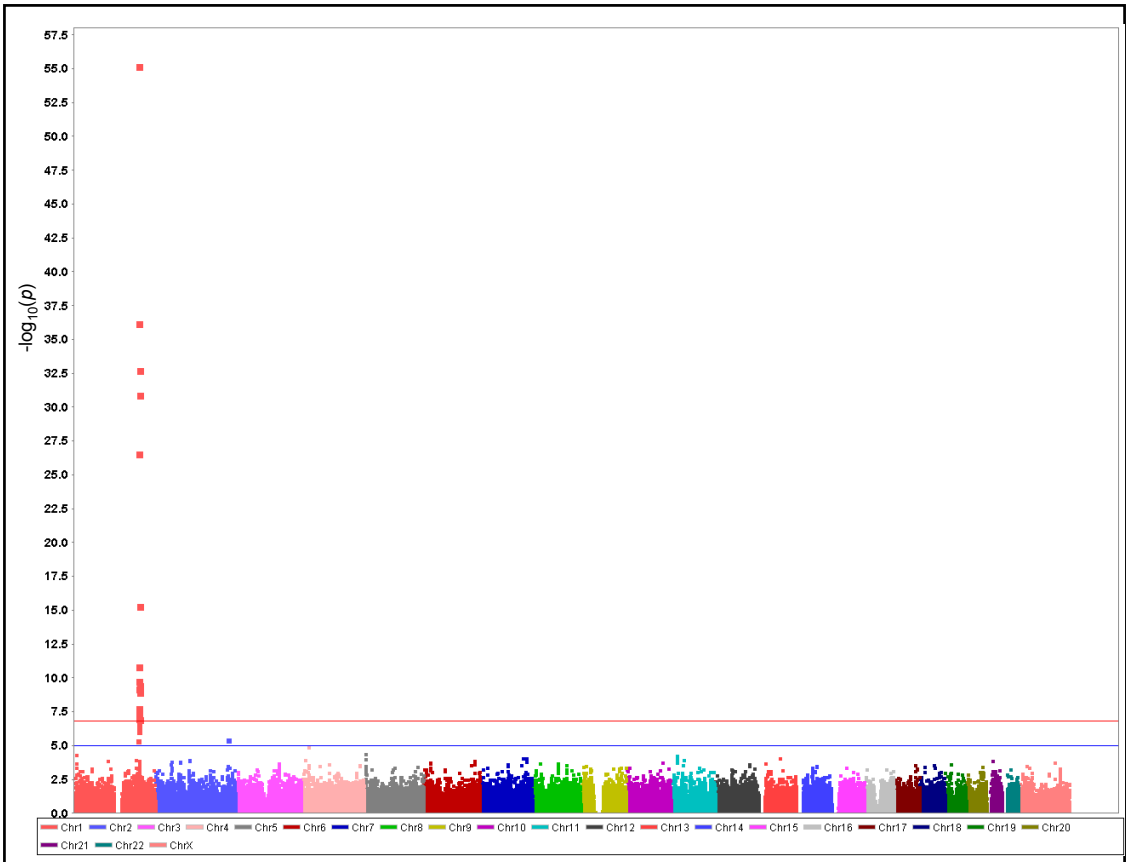


Figure 7.4 Results of the genome-wide quantitative trait locus association analysis of plasma CFH in the Vis population.

Log_{10} transformed significance values for plasma CFH test of association are presented for each chromosome, using the software Haploview. An association test of 308,144 markers was carried out in 941 individuals. Significant and suggestive thresholds of significance (after correction for multiple testing) are represented by red ($-log_{10}(P)$=6.8) and blue ($-log_{10}(P)$= 5.0) lines, respectively.

SNP	Minor allele	MAF	Chromosome	Position (bp)	Association Test (*p*)	HWD (*p*)	Genomic context
rs6677604	A	0.209	1q31	193418575	5.19E-56	0.674	*CFH* intron 11
rs2026429	C	0.051	1q31	193811328	2.67E-10	0.304	*ASPM* Intron 13
rs6003	G	0.049	1q31	193762678	7.95E-10	0.286	*F13B* His115Arg
rs1983298	G	0.088	2q35	219981339	3.39E-06	0.306	*PTPRN* intron 3
rs358248	T	0.371	4p15	22459983	1.19E-05	0.037	*GBA3* intron 4
rs7529369	T	0.077	1q31	192940547	1.39E-05	1.000	*KCNT2* intron 2
rs6555300	A	0.486	5p15	4587213	3.59E-05	0.439	Gene desert
rs950493	G	0.166	1p36	6879070	4.18E-05	0.816	*CAMTA1* intron 3
rs1487204	G	0.286	11p15	19552278	5.28E-05	0.937	139 kb upstream of *NAV2*

Table 7.3 Summary table of SNPs showing evidence of genome-wide association with plasma CFH.

Details of SNPs in chromosomal regions showing evidence of significant genome-wide associations ($P \leq$ 1.6E-07) and other loci showing suggestive associations ($P \leq$ 1.0E-05) or potential associations ($P \leq$ 1.0E-04) with plasma CFH concentration. Analysis was carried out in 941 individuals (related and unrelated). MAF: minor allele frequency, HWD: Hardy Weinberg disequilibrium. SNP positions are based on the NCBI Build 35 assembly.

Highly significant associations were observed between plasma CFH concentration and twelve polymorphisms within the Regulator of Complement Activation (RCA) cluster in chromosomal region 1q31, each with *P*-values meeting genome-wide significance ($P \leq$ 1.6E-07) after a Bonferroni correction for multiple testing (**Figure 7.4** and **Figure 7.5**). Nine of these were located at or near the *CFH* locus (**Figure 7.5**) and three were located in the *ASPM*, *F13B* and *KCNT2* genes, only the last of which is outside the RCA cluster (**Table 7.3**). The polymorphisms located in the *CFH* or *CFH*-related genes showed considerably greater genome-wide significance compared with those in the *ASPM*, *F13B* or *KCNT2* genes and were therefore thought to be better candidates for plasma CFH associations. All genotypes were consistent with HWE and had genotype call rates above 97%, which excluded the possibility of false-positive associations due to genotyping errors.

Five SNPs in or around the *CFH* gene showed suggestive genome-wide significance ($P \leq$ 1.0E-05) with plasma CFH (**Figure 7.5**). Similarly, five SNPs showing suggestive genome-wide associations with plasma CFH were found in chromosomal regions 2q35 (*PTPRN*), 4p15 (*GBA3*), 5p15 (intergenic region), 1p36 (*CAMTA1*) and 11p15 (intergenic region) (**Table 7.3**).

The most significant association with plasma CFH was obtained with SNP rs6677604 which is located in intron 11 of the *CFH* gene (P = 5.2E-56) (**Figure 7.4** and **Table 7.3**). Another SNP, rs2026429, which is located in intron 13 of the Abnormal Spindle-like Microcephaly-associated (*ASPM*) gene also showed a genome-wide significant association with plasma CFH (P = 2.7E-10) (**Table 7.3**). This SNP has a relatively low minor allele frequency (MAF) of 0.051 and may therefore be more likely to be a false-positive. However, the region of 1q31 shows unusually strong linkage disequilibrium (**Figure 7.5**). SNP rs6003, a coding variant of the *F13B* gene which encodes the coagulation Factor XIII B protein, was also found to be significantly associated with changes in plasma CFH concentration (P = 7.9E-10). Rs6003 is a non-synonymous coding SNP in *F13B* which is associated with a H115R substitution in the protein although it is less common in this sample set with a minor allele frequency of 0.049 (**Table 7.3**) and could therefore be a false-positive. SNP rs7529369 which lies upstream of *CFH* was also a significant hit from the genome-wide association analysis (P = 1.4E-08) (**Table 7.3**).

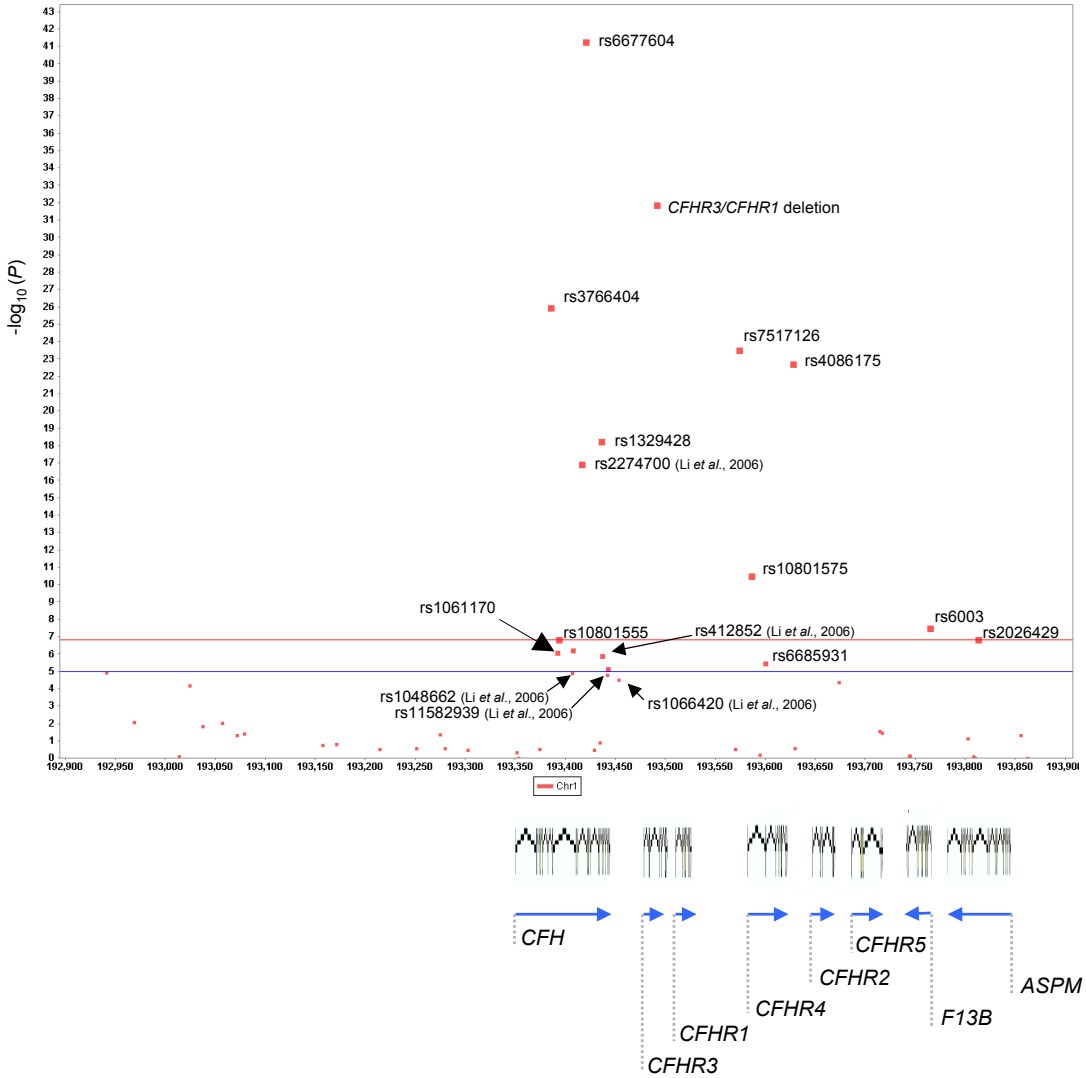
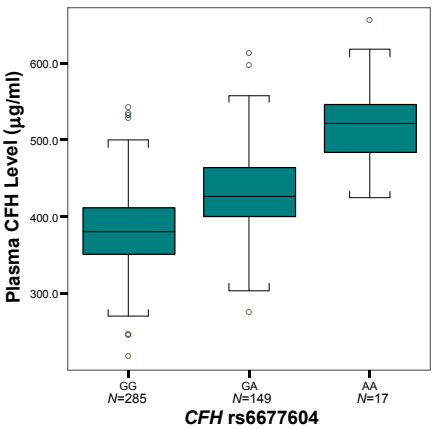


Figure 7.5 High resolution view of the Regulator of Complement Activation (RCA) cluster on 1q31 and SNPs associated with plasma CFH concentration in Vis.

Transformed *p*-values for the test of association in 472 unrelated individuals are shown. Thresholds for significant ($P \leq 1.6E\text{-}07$) and suggestive ($P \leq 1.0E\text{-}05$) genome-wide association are marked with red and blue lines respectively. The genomic context including the *CFH* gene cluster is also shown. Blue arrows represent direction of transcription.

Figure 7.6 Differences in plasma CFH concentration associated with rs6677604 genotypes.

Box plots showing differences in plasma CFH concentrations between the three genotype groups of *CFH* SNP rs6677604. The minor allele A is shown to be associated with an increase in concentration under an additive model. The bars represent the range from the smallest to the largest non-outlier measurement. Empty circles show the 5% extremes of the distribution. *N* shows the number of unrelated individuals in each group..



This SNP is located in intron 2 of the gene encoding a Potassium Channel protein of the subfamily T number 2 (*KCNT2*), which is expressed mainly in the human brain. SNP rs1983298 which lies in the third intron of the Protein Tyrosine Phosphatase Receptor type N (*PTPRN*) gene also showed suggestive genome-wide significance ($P \leq$ 1.0E-05) (**Table 7.3**). **Table 7.3** also includes a number of markers which showed weak associations but included several SNPs forming a peak with at least two SNPs showing *P*-values below 1.0E-04. SNP rs358248 is an intronic SNP in the Beta-Glucosidase Acid 3 (*GBA3*) gene with a minor allele frequency of 0.37 (P = 1.2E-05) (**Table 7.3**). *GBA3* encodes a cytosolic beta-Glucosidase enzyme which is predominantly expressed in the liver. SNP rs950493 is also an intronic SNP in the Calmodulin binding Transcription Activator 1 (*CAMTA1*) gene (P = 4.2E-05) (**Table 7.3**). The protein product acts as a transcription factor that interfaces with the calcium-calmodulin system of the cell to alter gene expression patterns and has been implicated in human episodic memory function (Huentelman *et al*., 2007). The remaining two SNPs, rs6555300 (P = 3.6E-05) and rs1487204 (P = 5.3E-05), both lie outside any known genes. SNP rs6555300 is located in a gene desert and rs187204 is 139 kb from the nearest gene (**Table 7.3**).

Eleven SNPs showing genome-wide significant associations with plasma CFH ($P \leq$ 1.6E-07) spanned a region of about 300 kb containing the *CFH* gene cluster including four of the *CFH*-related genes (*CFHR1-5*) (**Figure 7.5** and **Table 7.4**). The strongest association was with SNP rs6677604 which is located in intron 11 of *CFH* and explains about 28.3% of the phenotypic variance after adjusting for fixed effects and covariates (P = 1.4E-36) (**Table 7.4**). The rs6677604 A (minor) allele was associated with a 59.4 μg/ml increase in mean plasma CFH level under an additive model. Less common polymorphisms identified in *F13B* and *ASPM* genes were also associated with large changes in mean plasma CFH concentration reaching genome-wide significance due to their low allele frequencies (**Table 7.4**). Differences in plasma CFH concentration associated with genotypes of the *CFH* SNP rs6677604 are presented in box-plots in **Figure 7.6.** The diagram shows a progressive increase in plasma CFH concentration associated with each copy of rs6677604 allele A, with the highest levels being associated with the AA genotype.

SNP	Minor Allele	MAF	Position	Genomic Context	β (S.E.) μg/ml	R^2	*P* (Association Test)	*N*
rs6677604	A	0.201	194953541	*CFH* intron 11	59.4 (4.4)	0.283	1.41E-36*	942
CFHR3/CFHR1 deletion	D	0.197	194988783 - 195075128	86.3 kb deletion of the *CFHR3/CFHR1* genes	55.1 (4.6)	0.235	3.21E-27*	930
rs3766404	C	0.133	194918455	*CFH* intron 6	58.3 (5.4)	0.196	2.87E-21*	944
rs7517126	G	0.124	195106895	16.9 kb upstream of *CFHR4*	55.8 (5.5)	0.180	7.89E-19*	948
rs4086175	G	0.121	195160949	18.6 kb upstream of *CFHR2*	56.3 (5.6)	0.177	4.54E-18*	934
rs1329428	A	0.421	194969433	*CFH* intron 15	34.9 (3.9)	0.143	1.34E-13*	956
rs2274700	T	0.421	194949570	*CFH* (A473A) exon 10	33.6 (3.9)	0.137	2.83E-12*	934
rs10801575	T	0.309	195119404	4.4 kb upstream of *CFHR4*	28.2 (4.2)	0.087	7.41E-06*	940
rs6003	G	0.044	195297644	*F13B* (H115R) exon 3	52.0 (9.3)	0.062	0.0077*	942
rs2026429	C	0.047	195346294	*ASPM* intron 13	48.2 (9.1)	0.056	0.0357*	950
rs10801555	A	0.383	194926884	*CFH* intron 9	-22.2 (4.2)	0.057	0.0372*	930
rs2019724	A	0.427	193406574	*CFH* intron 9	-20.5 (4.1)	0.050	7.56E-07	956
rs1061170	C	0.388	193390894	*CFH* (Y402H) exon 9	-20.8 (4.2)	0.050	1.08E-06	924
rs412852	C	0.426	193435364	*CFH* intron 15	-19.9 (4.1)	0.048	1.57E-06	940
rs6685931	C	0.399	193598890	*CFHR4* intron 1	-18.9 (4.1)	0.045	3.79E-06	932
rs1065489	T	0.154	193441431	*CFH* intron 18	-24.2 (5.4)	0.041	8.46E-06	950

* *P*-values after a Bonferroni correction for multiple testing.

Table 7.4 Summary of the polymorphisms showing full and suggestive associations with plasma CFH after correction for multiple testing.

List of polymorphisms found to be associated with plasma CFH concentration in a genome-wide association analysis after correcting for multiple testing and covariate effects. β is the regression coefficient for plasma CFH concentration on genotype which indicates the direction and size of additive effect for the minor allele. R^2 shows the proportion of the adjusted phenotypic variance explained by each marker, independent of the effects of sex, age-by-sex interaction and BMI. Suggestive, unadjusted associations are shown under the dashed line. *N* is the number of chromosomes of unrelated individuals used in the analysis. MAF: minor allele frequency. Positions are based on the NCBI Build 35 assembly.

Three of the remaining eight genome-wide significant associations occurred outside the *CFH* gene (**Figure 7.5**). These included the deletion of the *CFHR3/CFHR1* genes (P = 3.2E-27) and two SNPs rs7517126 (P = 7.9E-19) and rs4086175 (P = 4.5E-18) which are upstream of the *CFHR4* and *CFHR2* genes, respectively (**Table 7.4**). *CFH* rs2274700 was found to be the only coding variant showing significant genome-wide association (P = 2.8E-12), explaining about 14% of the phenotypic variance and encoding an Ala473Ala synonymous substitution within *CFH* exon 10 (**Table 7.4**). Five other SNPs, including the SNP rs1061170 (*CFH* Y402H) reached suggestive genome-wide significance for association ($P \leq$ 1.0E-05) (**Figure 7.5** and **Table 7.4**).

7.2.2.3 Analysis of linkage disequilibrium across the *CFH* locus

Twenty-nine genotyped markers spanning the *CFH* and *CFH*-related genes were used to assess the extent of pairwise linkage disequilibrium (LD) in the region. The analysis was carried out using the software Haploview® (Barrett *et al*., 2005) in 492 unrelated individuals from the Vis sample set (**Figure 7.7**). LD was measured in terms of r^2 so that the relationship between markers was not affected by differences in allele frequencies, as explained in Section 2.5.10. Five haplotype blocks were constructed, based on a solid spine of LD and SNP haplotypes with a frequency greater than 1% are shown in **Figure 7.7**. Pairwise LD between SNP rs6677604 and polymorphisms reported in the literature to affect susceptibility to AMD including the *CFHR3/CFHR1* deletion (Hughes *et al*., 2006), SNP rs1061170 (*CFH* Y402H) (Haines *et al*., 2005; Edwards *et al*., 2005; Klein *et al.*, 2005, Zareparsi *et al*., 2005a) and *CFH* rs2274700 (Li *et al*., 2006) was specifically assessed. The *CFHR3/CFHR1* deletion was found to be in strong LD with *CFH* rs6677604 (r^2 = 0.84) and occurred in the same haplotype block. The most common haplotype harbouring the rs6677604 minor (A) allele and the *CFHR3/CFHR1* deleted allele was found to have a frequency of 0.12 in the Vis sample. This haplotype also contained the *CFH* rs1061170 T (Tyr^{402}) allele, previously shown to have a protective effect on AMD. Pairwise LD was found to be relatively weak ($r^2 < 0.35$) between rs6677604 and either rs2274700 or rs1061170.

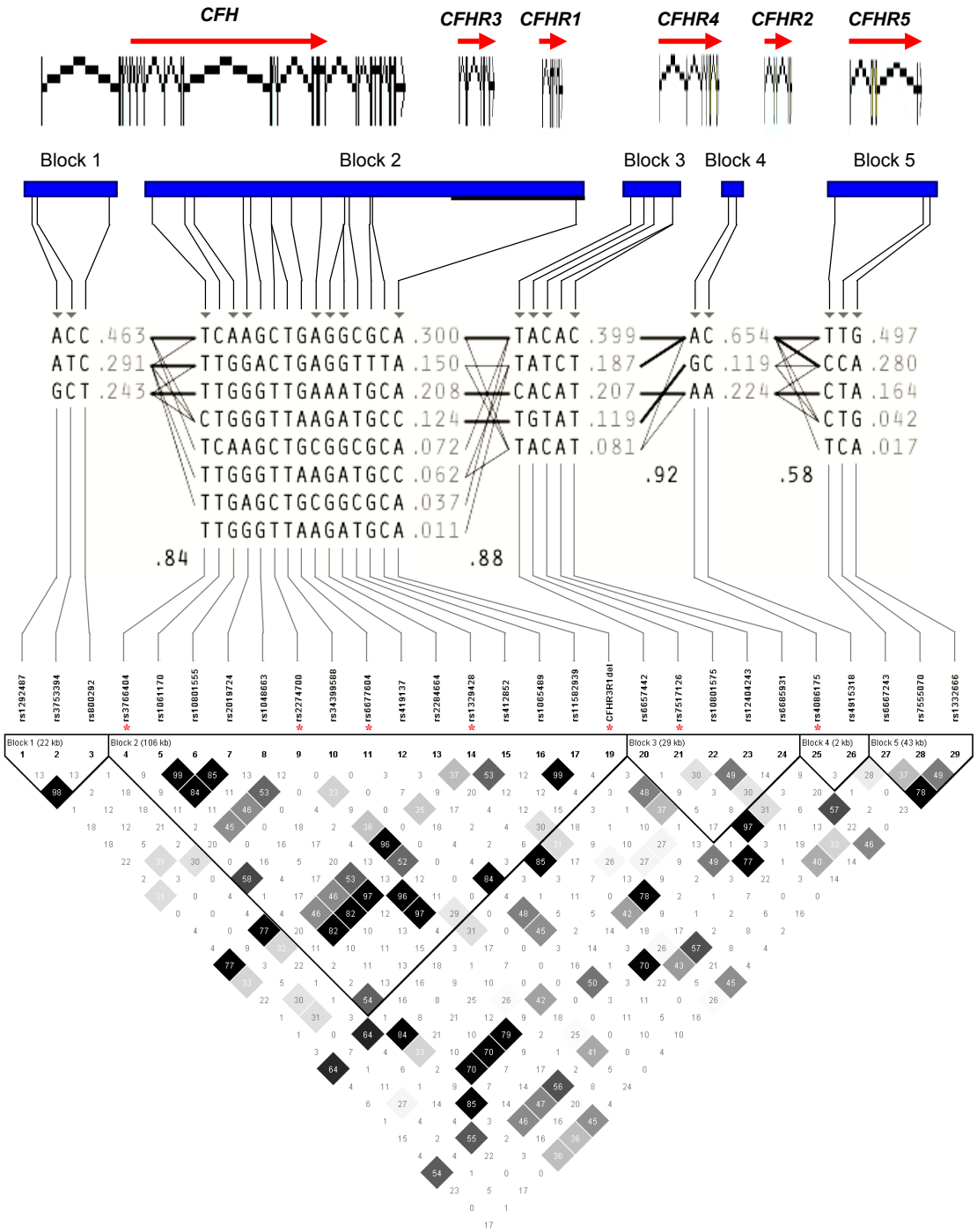


Figure 7.7 Analysis of linkage disequilibrium across the *CFH* locus based on the analysis of 492 unrelated individuals from Vis island.

Pairwise LD is shown in terms of the squared correlation coefficient (r^2) between 28 markers spanning *CFH* and *CFH*-related genes. Five haplotype blocks were constructed based on a solid spine of LD and the tagging markers for each haplotype are marked with grey vertical arrow heads. Inter-block LD is also presented. The frequency of each haplotype is shown next to it in grey. *CFH* genomic regions and the position of markers are shown at the top. Markers showing significant genome-wide association are marked with red asterisks. Black squares represent high LD (r^2>0.70) with fading colours showing reduction in LD. The level of pairwise LD is shown inside each square. Alleles of the *CFHR3/CFHR1* deletion are encoded as C, deleted and A, non-deleted. Analysis was carried out using the software Haploview.

Figure 7.7 also shows the seven markers showing genome-wide significant associations with plasma CFH concentration (marked by red asterisks), represented in three haplotype blocks with inter-block LD (r^2) > 0.80. Five of the markers occurred in haplotype block 2. A further haplotype analysis was carried out firstly to assess whether the variance observed in plasma CFH was better explained by *CFH* rs6677604 alone, or by a haplotype containing this SNP. Secondly, a haplotype analysis was carried out to assess whether the effect of rs6677604 on plasma CFH was independent from that of other variants found in the genome-wide association analysis to be significantly associated with plasma CFH.

7.2.2.4 Identification of a multi-marker model for plasma CFH concentration

Multi-marker analysis of plasma CFH was carried out, firstly using the "backward" stepwise linear regression methodology implemented in the statistical software SPSS v.12. Analysis of "backward" regression was based on the probability of $P > 0.05$ of the model's F-ratio as a removal criterion of a component. The F-ratio is calculated by dividing the average improvement in prediction by the model (MS_M) by the average difference between the model and the observed data (MS_R). All SNPs as well as the fixed effect sex and covariates age and BMI were initially included in the model and the contribution of each predictor was assessed based on the significance value of the t-test for that predictor. The significance value was checked against the removal criterion for F ($P > 0.05$) and if a predictor (SNP or covariate) was not making a statistically significant contribution to how well the model predicted the outcome variable it was removed and the model was re-estimated for the remaining predictors. The contribution of the remaining model components was then re-assessed.

An important assumption in linear regression analysis is the normality of the distribution. Plasma CFH distribution was checked for normality using the Kolmogorov-Smirnov test. The test statistic with the null hypothesis of a non-normal distribution was found to be non-significant at the 1% level ($P = 0.01$).

494 unrelated individuals from the island of Vis were included in the analysis and 22 were removed due to low genotyping call rates (< 90%). Twenty-seven bi-

allelic markers, spanning the *CFH* and *CFH*-related genes, were included in the multi-marker analysis as covariates in a model assuming an additive effect of the minor allele (1 d.f.). Age and BMI were included as covariates. Sex was analysed as a fixed effect.

Twenty models with different combinations of predictors were assessed. The best model which explained the highest proportion of the phenotypic variance, consisted of six SNPs, age and BMI (**Table 7.5**). Age and BMI were found to significantly correlate with plasma CFH concentration in this model ($P < 0.05$). The most strongly associated SNP, rs6677604, was also included in this model. The marker data was then checked for collinearity, which exists when there is a strong correlation between two or more predictors in the regression model (Section 2.5.7.2.1). Two parameters were used to estimate collinearity between predictors, the variance inflation factor (VIF) and tolerance (1/VIF). VIF indicates whether a predictor has a strong linear relationship with the other predictor(s). VIF values greater than 10 are taken as indicative of redundant predictors (Myers, 1990). Similarly, values of the tolerance statistic below 0.1 are indicative of collinearity (Menard, 1995). Markers rs3766404 and rs7517126 were found to contain redundant information (VIF > 10). This was also evident from the high level of pairwise LD between these two SNPs ($r^2 = 0.84$) (**Figure 7.7**). However, removal of SNP rs3766404 resulted in a reduction of the VIF below 10 indicating that the effect associated with SNP rs3766404 was now fully explained by SNP rs7517126. Furthermore, the effect of SNP rs3766404 would be accounted for by the high pairwise LD between rs3766404 and rs7517126 ($r^2 = 0.85$) and to a lesser extent between rs3766404 and rs6677604 ($r^2 = 0.58$).

A univariate analysis of variance approach was used to validate the above model and to evaluate the proportion of the phenotypic variance accounted for. Analysis was carried out as part of the general linear model (GLM) methodology implemented in the software SPSS v.12. Five SNPs, rs1292487, rs6677604, rs7517126, rs10801575 and rs6667243, were used in a test of between-subject effects, under an additive model (1 d.f.), including age and BMI as significant covariates (**Table 7.6**).

					Collinearity Statistics	
	β (μg/ml)	S.E.	t-test	*P*	Tolerance	VIF
Age	0.54	0.19	2.87	4.4E-03	0.956	1.05
BMI	2.96	0.63	4.68	4.2E-06	0.957	1.05
rs1292487	17.02	7.30	2.33	2.0E-02	0.341	2.93
rs3766404	-41.88	19.80	-2.12	3.5E-02	0.082	12.24
rs6677604	61.96	9.54	6.49	3.2E-10	0.253	3.95
rs7517126	58.40	17.79	3.28	1.1E-03	0.097	10.30
rs10801575	-27.40	8.02	-3.42	7.2E-04	0.230	4.35
rs6667243	14.77	7.07	2.09	3.7E-02	0.273	3.67

Table 7.5 Results of the linear regression test for collinearity of markers used in the analysis of plasma CFH.

27 markers spanning *CFH* and the *CFH*-related genes together with age, sex and BMI were used in a "backward" linear regression analysis. The table shows the model which explained the highest proportion of the plasma CFH phenotypic variance. A test of collinearity was also carried out to detect SNPs which carried redundant information. CFH rs3766404 was removed from further analysis due to its high variance inflation factor (VIF >10). β shows the changes in mean plasma CFH concentration under an additive model of the minor allele for each SNP. t-statistics and their significance values (*P*) are also shown. VIF (variance inflation factor) and tolerance are explained in the text. Regression analysis was carried out using the statistical software SPSS v.12.0.

	Type III Sum of Squares	d.f.	Mean Square	F	*P*	R^2
rs1292487	6159.4	1	6159.4	2.51	1.1E-01	0.006
rs6677604	170912.8	1	170912.8	69.73	9.6E-16	0.140
rs10801575	14109.5	1	14109.5	5.76	1.7E-02	0.013
Age	27548.7	1	27548.7	11.24	8.7E-04	0.026
BMI	71042.7	1	71042.7	28.99	1.2E-07	0.064

R^2_{model} = 0.380 (Adjusted R^2_{model} = 0.369)

Table 7.6 Results of the univariate analysis of plasma CFH using a multi-marker model obtained by 'backward' regression analysis.

Non-redundant, significant markers from the 'backward' linear regression analysis were tested in a multi-marker analysis of plasma CFH concentration in the Croatian sample set. R^2 represents the proportion of the phenotypic variance explained by each component. R^2_{model} is a measure of how much of the variability in the trait distribution is accounted for by the model. For adjusted R^2_{model} see text. F-statistics and their significance value (*P*) are also shown. Regression analysis was carried out using the statistical software SPSS v.12.0.

All SNPs in the 5-marker model excluding rs6667243 ($P > 0.05$) and rs7517126 ($P > 0.05$) were found to be significantly associated with plasma CFH concentration. The best model therefore consisted of the SNP rs1292487, located 1.6 kb upstream of *CFH*; SNP rs6677604, in intron 11of the *CFH* gene and SNP rs10801575 located 4.4 kb upstream of *CFHR4*, and the covariates age and BMI.

The proportion of phenotypic variance explained by the full model components was estimated at 38% (R^2_{model}=0.38) (**Table 7.6**). The adjusted R^2 accounts for the model complexity and shows that 36.9% of the trait variance would be accounted for if the model had been derived from the population from which the sample was taken. Therefore, the difference of 1.1% between the R^2_{model} and the adjusted R^2_{model} shows that if the model were derived from the population rather than a sample it would account for approximately 1.1% less variance in the outcome. This means that the model was a good predictor of the correlation between model components and the trait distribution at the population level. When the three SNPs were included in the model, without the covariates, the adjusted R^2 value was estimated to be 29.2% which did not significantly differ from the effect of the single SNP ($R^2_{rs6677604}$ = 28.3% (**Table 7.4**)). The marker data was checked for all significant 2-way interactions using the same methodology and none were detected. The three remaining SNPs in the best model, rs1292487, rs6677604 and rs10801575, were then used in a conditional haplotype analysis of plasma CFH concentration with age and BMI as covariates, which is described below.

7.2.2.5 Haplotype-based association analysis of CFH

The 3-SNP model identified by 'backward' regression was used in a haplotype association analysis using the conditional haplotype-based testing tool, implemented in the software package PLINK (Purcell *et al*., 2007). The underlying theory of haplotype analysis in PLINK has been described in the Section 2.5.7.3.

Six common haplotypes (minor haplotype frequency (MHF) ≥ 0.01), out of a possible total of 8, were identified in 410 unrelated individuals from Vis (F_{model} = 28.9, 5 d.f., P = 5.5E-25) (**Table 7.7**).

rs1292487	rs6677604	rs10801575	Frequency	β (μg/ml)	*P* (HST)	*P* (HST | Y402H)	*P* (HST | deletion)
T	**A**	**T**	0.177	55.51	1.1E-26	2.3E-20	5.3E-03
C	G	T	0.016	-63.99	0.083	5.5E-03	0.345
T	G	T	0.105	-65.58	1.5E-03	6.2E-06	0.038
T	G	C	0.481	-58.11	2.1E-09	2.3E-03	0.088
C	G	C	0.216	-48.50	0.290	1.2E-03	0.092
T	**A**	**C**	0.013	16.24	2.9E-04	3.1E-04	0.300

Table 7.7 Association analysis of the 3-SNP haplotypes with plasma CFH concentration in the Vis population.

Haplotype association analysis using the 3 SNPs selected from the 'backward' regression analysis was carried out in 410 unrelated individuals from Vis. Six common haplotypes (minor haplotype frequency (MHF) > 0.01) were identified and haplotype specific tests (HST) for the effect associated with each haplotype relative to the remaining haplotypes in the model were estimated (β). Haplotypes associated with an increase in CFH levels are highlighted in bold. Haplotype specific tests, conditional on *CFH* Y402H and the *CFHR3/CFHR1* deletion polymorphisms were also carried out. Analysis was carried out using the conditional haplotype association tool implemented in the software package PLINK (Purcell *et al*., 2007).

Age and BMI were included as covariates in the analysis. Sixty-two individuals were removed due to low genotyping call rate (< 90%).

Two out of the six haplotypes identified using SNPs rs1292487, rs6677604 and rs10801575 were associated with an increase in mean plasma CFH concentration (β_{TAT} = 55.51 μg/ml, *P* = 1.1E-26; β_{TAC} = 16.24 μg/ml, *P* = 2.9E-04) (**Table 7.7**). Four out of the six haplotypes were associated with reductions in mean plasma CFH concentration (**Table 7.7**). However, only TGT (β = -65.58 μg/ml, *P* = 1.5E-03) and TGC (β=-58.11 μg/ml, *P*=2.1E-09) resulted in a significantly decreased plasma CFH.

The effect of each common haplotype on plasma CFH concentration, conditional on the *CFH* Y402H and *CFHR3/CFHR1* deletion polymorphisms, was analysed using the conditional haplotype association test implemented in the software package PLINK v1.01 (Purcell *et al*., 2007). Conditional association between plasma CFH and the TAT haplotype, given the rs1061170 (*CFH* Y402H) variant, showed that the effect associated with the haplotype was independent of rs1061170 (*P* = 2.3E-23) (**Table 7.7**). In other words, the reduced associated with the TAT haplotype was still highly significant in the absence of the rs1061170 (*CFH* Y402H) effect. The decrease observed in the significance value can be explained by a reduction in the degrees of freedom of the model.

The most common haplotype, TGC (MHF = 0.481) was associated with a decrease of 58.1 μg/ml in plasma CFH concentration and was also found to have an independent effect (*P* = 2.3E-03) from that of *CFH* rs1061170 (Y402H) (**Table 7.7**).

A test of association between the TAT haplotype and plasma CFH, conditional on the *CFHR3/CFHR1* deletion polymorphism resulted in a sharp fall in the *P*-value (*P* = 5.3E-03), however the haplotype effect on plasma CFH remained significant even when conditioned on the absence of the deletion polymorphism (**Table 7.7**) despite a high level of pairwise LD between the two polymorphisms (r^2 = 0.84) (**Figure 7.7**). In contrast, the effect associated with the TGC haplotype was non-significant when conditioned on the deletion polymorphism, indicating that the haplotype effect is not independent (*P* = 0.088) (**Table 7.7**). Some of the decrease in *P*-value could however be the result of reduction in the number of degrees of freedom in the test. The TGT haplotype however, remained significant when conditioned on the deletion polymorphism (*P* = 0.038).

7.2.3 QTL analysis of plasma CFH in the Dutch sample set

Plasma CFH concentrations were measured in 500 unrelated individuals from a Dutch isolated (Rucphen) population, and used as a replication series to validate the findings of the QTL association analysis from Vis. Markers were therefore genotyped as described in the next section. Information on age and sex of the individual samples was also obtained, as shown in **Table 2.1**.

7.2.3.1 Genotyping of markers

A total of eight markers were selected to be genotyped in the Dutch samples. These included the most significant SNP associations from the genome-wide association analysis, the *CFHR3/CFHR1* deletion polymorphism and rs1061170 (*CFH* Y402H). Attempts to design a genotyping assay for rs4086175 were unsuccessful and this SNP was consequently excluded from the analysis. *CFH* SNP rs2274700 was genotyped by direct sequencing.

7.2.3.2 Single marker association analysis of plasma CFH

Single marker association with plasma CFH was carried out using the linear regression methodology implemented in the genome-wide association software package PLINK v.0.99s. The results confirmed the findings of the Vis population sample set since *CFH* SNP rs6677604 was the most significantly associated marker (P = 4.9E-25) (**Figure 7.8**). *CFH* rs1061170 was also found to be significantly associated with differences in plasma CFH concentration in this sample set (P = 1.8E-04) (**Table 7.8**).

The effect of individual markers associated with plasma CFH concentration in the Dutch sample set was assessed in a univariate analysis of variance. Age was used as a significant covariate for plasma CFH in all analyses (**Table 7.8**). *CFH* SNP rs6677604 was found to explain 21.8% of the phenotypic variance (P = 4.9E-25). Allele frequencies of the markers were generally similar between the two population samples except for the *CFHR3/CFHR1* deletion polymorphism (MAF_{Vis} = 0.197 compared with $MAF_{Rucphen}$ = 0.295) (**Table 7.8**).

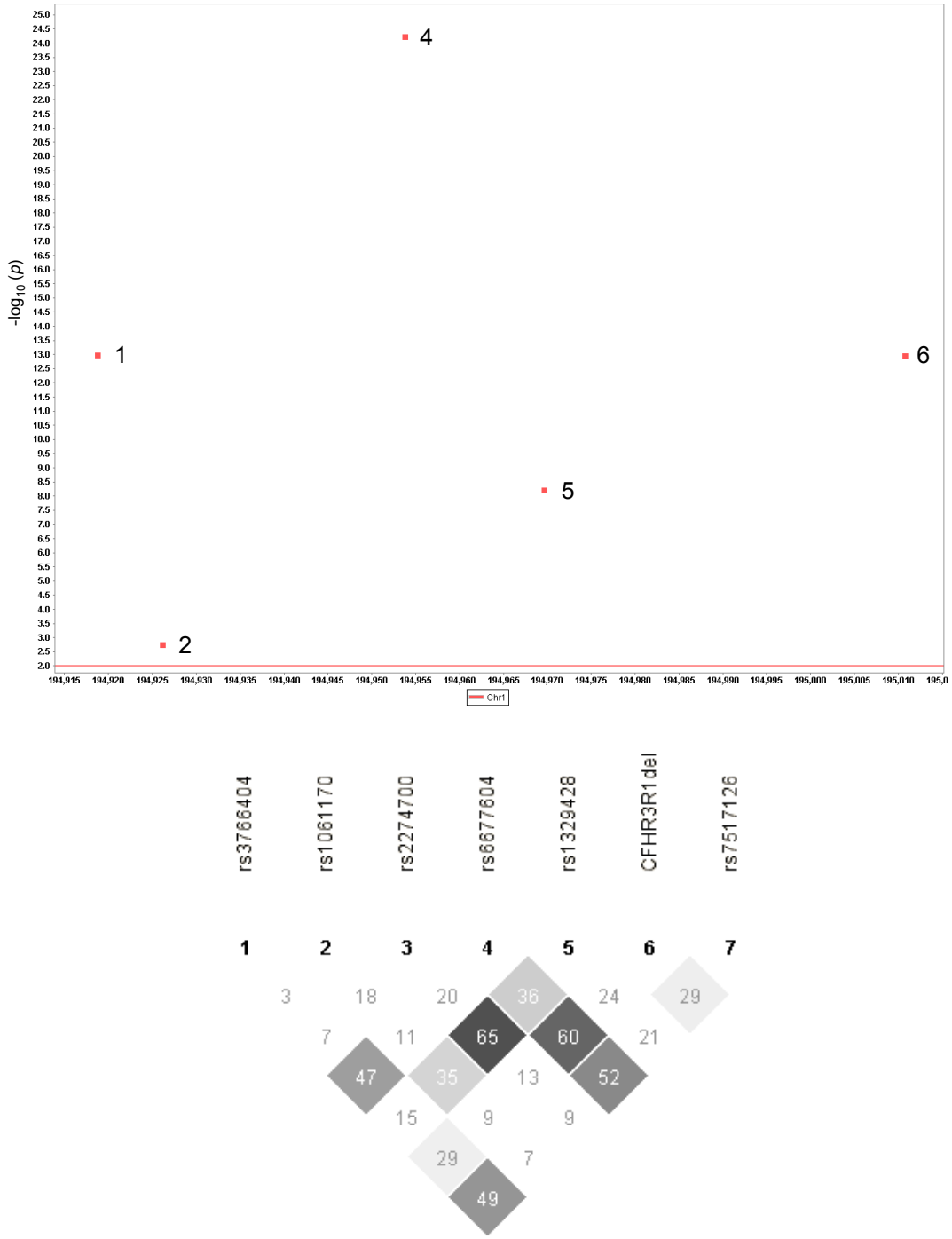


Figure 7.8 Plasma CFH association test and linkage disequilibrium analysis in the Dutch (Rucphen) isolate.

Results of the association analysis of plasma CFH concentration in 453 unrelated Dutch individuals. The significance threshold after correction for multiple testing using the Bonferroni correction, is marked by the red line. Two markers rs2274700 and rs7517126 were excluded from the analysis due to low genotyping call rates (< 90%). Pairwise linkage disequilibrium (LD) between each of the seven markers is shown in terms of the squared correlation (r^2). The level of pairwise LD is shown inside each square box.

Marker ID	Minor Allele	MAF	R^2	*p*	*N*
rs6677604	A	0.211	0.218	4.90E-25	896
CFHR3/CFHR1 deletion	D	0.295	0.132	1.26E-14	866
rs3766404	C	0.167	0.113	5.01E-13	898
rs1329428	T	0.453	0.091	2.75E-10	860
rs7517126	G	0.132	0.101	1.05E-09	722
rs2274700	T	0.434	0.086	6.07E-06	470
rs1061170	C	0.322	0.032	1.78E-04	898

Table 7.8 Single marker associations with plasma CFH concentration in the Dutch (Rucphen) isolate.

Results of single marker univariate regression analysis in unrelated Dutch individuals. All analyses included age as a significant covariate for plasma CFH levels. Minor allele frequencies (MAF) are shown for each marker. R^2 represents the proportion of phenotypic variance explained by each marker and the significance value (*P*) for the association. *N* is the number of chromosomes of unrelated individuals analysed.

The SNP rs6677604 was found to explain a slightly smaller proportion of the variance in plasma CFH in the Rucphen population (R^2 = 21.8%) compared with Vis (R^2 = 28.3%) despite having the same minor allele frequency in both populations (MAF = 0.21).

Analysis of linkage disequilibrium across the seven markers in the Dutch samples revealed lower levels of pairwise LD than in the Vis sample set. *CFH* SNPs rs2274700 and rs1329428 (r^2 = 0.65), and rs6677604 and *CFHR3/CFHR1* deletion (r^2 = 0.60), were the only pairs of markers with detectable LD.

7.2.4 Analysis of *CFH* association with AMD

Association analysis was carried out in 382 cases of age-related macular degeneration (AMD) and 201 sex and age-matched disease-free controls. The average age of individuals in the series was 75.6 years (± 9.5 SD) in cases and 73.1 years (± 8.0 SD) in controls (**Table 2.1**). All individuals were unrelated and of Scottish ethnicity. The criteria used for assessing disease status and AMD grading were described in Section 2.2.

Twenty-five SNPs, including those from other genes previously found to influence susceptibility to AMD had previously been genotyped in all samples. Seven new markers including six SNPs and the *CFHR3/CFHR1* deletion polymorphism were added to the genotype data. Details of allele frequencies and the results of the case-control association analyses are summarised in **Table 7.9**. Using a Fisher's exact test of association, the association showing the highest significance was observed with the rs1061170 (*CFH* Y402H) variant (OR = 2.23, CI: 1.73 - 2.90, *P* = 7.1E-10) which explained 6.8% of the AMD disease variance estimated by the correlation coefficient of the logistic regression η^2 (**Table 7.9**, **Figure 7.9**). Five markers including *CFH* rs3766404 (OR = 0.49, CI: 0.33 - 0.73, *P* = 3.9E-04; η^2 = 0.025), rs2274700 (OR = 0.55, CI: 0.42 - 0.73, *P* = 3.3E-05; η^2 = 0.034), rs6677604 (OR = 0.53, CI: 0.38 - 0.75, *P* = 4.6E-04; η^2 = 0.022), *CFHR3/CFHR1* deletion (OR = 0.57, CI: 0.41 - 0.81, *P* = 1.4E-03; η^2 = 0.017) and rs7517126 (OR = 0.51, CI: 0.34 - 0.76, *P* = 1.3E-03; η^2 = 0.023) were all found to exhibit significantly protective effects against AMD in single SNP tests of association (**Table 7.9**).

					Minor allele Frequency Comparisons				Genotypic model	
Marker ID	**Minor allele**	**Position**	**Context**	η^2	**OR (95% C.I.)**	***P***	**F_{cases}**	**$F_{Controls}$**	***P* (2d.f.)**	***P*$_{(HWD)}$**
rs1061170	C	194925860	*CFH* (Y402H) exon 9	0.068	2.23 (1.73-2.90)	7.14E-10	0.541	0.345	4.53E-08	0.5496
rs1329428	A	194969433	*CFH* intron 15	0.058	0.45 (0.35-0.59)	7.56E-09	0.247	0.419	6.90E-07	0.1898
rs412852	C	194970330	*CFH* intron 15	0.053	2.04 (1.58-2.63)	3.92E-08	0.582	0.406	7.71E-07	0.1237
rs2274700 *	T	194949570	*CFH* (A473A) exon 10	0.034	0.55 (0.42-0.73)	3.33E-05	0.253	0.379	1.68E-03	0.1420
rs3766404	C	194918455	*CFH* intron 6	0.025	0.49 (0.33-0.73)	3.98E-04	0.081	0.152	1.84E-02	1.0000
rs6677604	A	194953541	*CFH* intron 11	0.022	0.53 (0.38-0.75)	4.61E-04	0.114	0.194	3.38E-02	0.0307
rs7517126	G	195106895	16.9 kb upstream of *CFHR4*	0.023	0.51 (0.34-0.76)	1.33E-03	0.079	0.144	4.12E-02	0.8113
CFHR3/CFHR1 deletion	D	194988783 - 195075128	86.3 kb deletion of *CFHR3/CFHR1* genes	0.017	0.57 (0.41-0.81)	1.37E-03	0.129	0.205	7.27E-02	0.1823
rs4086175 *	G	195160949	18.6 kb upstream of *CFHR2*	0.017	0.58 (0.40-0.87)	8.68E-03	0.087	0.139	1.74E-01	1.0000
rs1048663 *	A	194941605	*CFH* intron 9	0.001	0.91 (0.66-1.27)	6.10E-01	0.164	0.177	3.10E-01	0.0130
rs11582939 *	T	194976780	*CFH* intron 18	0.001	0.92 (0.66-1.28)	6.71E-01	0.165	0.177	4.17E-01	0.0309
rs1066420 *	G	194987678	4.4 kb downstream of *CFH*	0.000	0.94 (0.66-1.35)	7.85E-01	0.147	0.154	-	0.0000
rs10490924	T	124204438	6.6 kb upstream of *ARMS2*	0.064	2.61 (1.93-3.53)	9.65E-11	0.3806	0.1906	3.40E-07	0.0005
rs11200638	A	124210534	*HTRA1* promoter	0.058	2.51 (1.86-3.39)	5.57E-10	0.3743	0.1923	8.42E-07	0.0014

Table 7.9 AMD case-control association analysis in the Scottish cohort.

The results of Fisher's exact test of association in 382 unrelated cases of age-related macular degeneration and 201 sex and age-matched controls. Six biallelic markers showing significant genome-wide association with plasma CFH were included. Five SNPs taken from Li *et al*. (2006) are marked with asterisks. SNPs rs10490924 and rs11200638 from chromosomal region 10q26 shown to be associated with AMD are also included. Positions are based on March 2006 genome assembly. η^2 shows the proportion of the disease variance explained by each polymorphism; OR, odds ratios calculated for the effect of the minor allele; C.I., confidence interval; F, minor allele frequency; $P_{(HWD)}$, significance value for deviations from Hardy-Weinberg equilibrium.

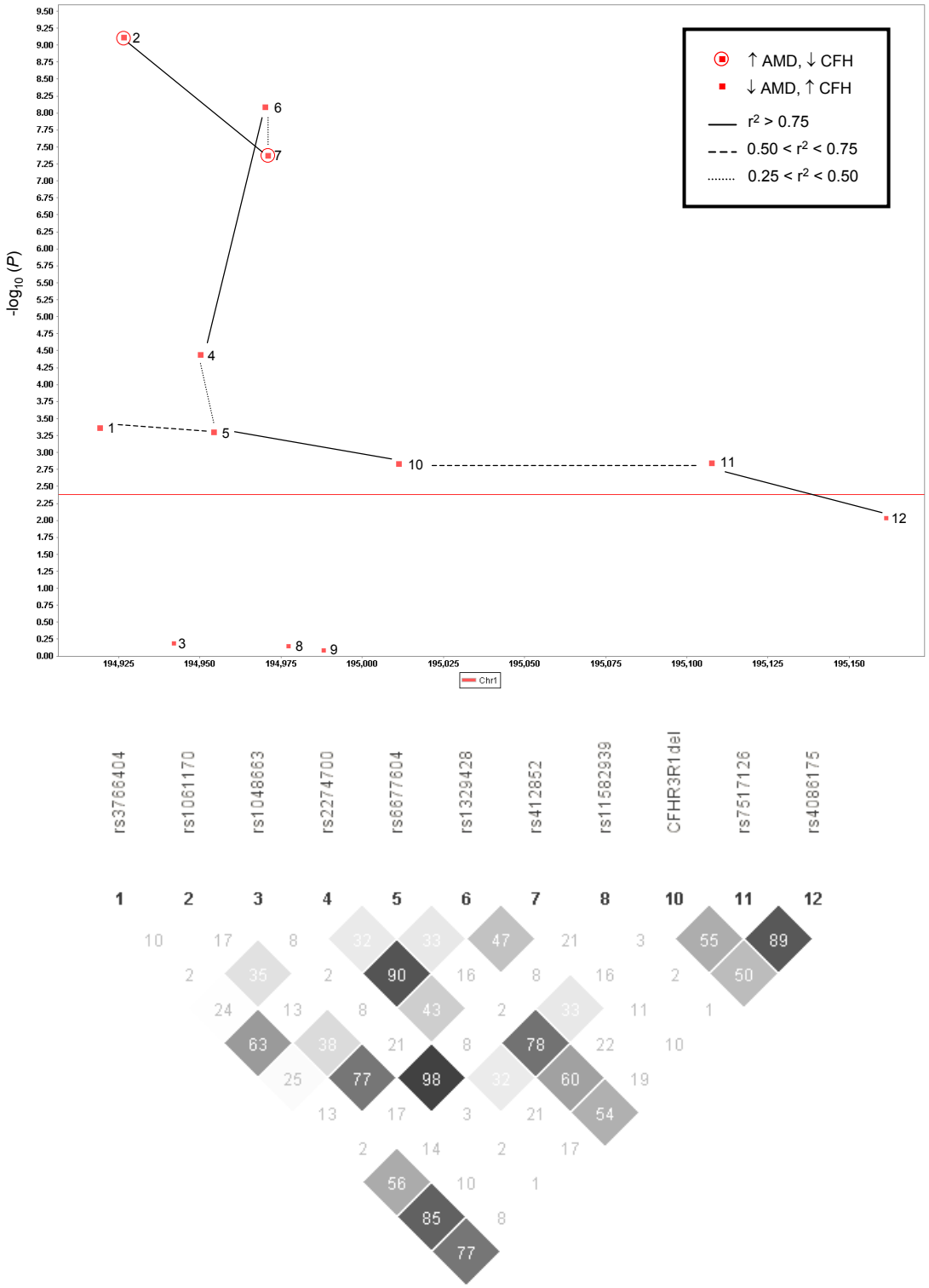


Figure 7.9 Test of association in unrelated cases of AMD and controls.

Plot of $\log_{10}$-transformed *P*-values for Fisher's exact test of single-marker association using 382 unrelated cases of AMD and 201 disease-free controls are shown. The significance threshold after Bonferroni correction for multiple testing is marked with a red line. Linkage disequilibrium (LD) between markers used in the analysis is shown below, and on the association plot, as pairwise squared correlations (r^2). SNP 9 (rs1066420) is excluded from the LD map due to deviation from HWE.

Analysis of linkage disequilibrium in the Scottish samples showed that associated SNPs fell into three groups of LD ($r^2 > 0.70$) (**Figure 7.9**). The first group contained *CFH* SNPs rs1061170 (SNP # 2 in Figure 7.9) and rs412852 (SNP # 7), both significantly increasing the risk of AMD with ORs of 2.23 (95% CI 1.7 - 2.9) and 2.04 (95% CI 1.6 - 2.6), respectively. The second group consisted of *CFH* rs1329428 (SNP # 6) and rs2274700 (SNP # 4) both showing a protective effect against AMD. *CFH* rs1329428 (*P* = 7.6E-09) was found to be more significantly associated with AMD than rs2274700 (*P* = 3.3E-05). The third LD group consisted of a number of markers with lower but still significant effects on disease predisposition. These included *CFH* rs3766404 (SNP # 1), rs6677604 (SNP # 5), *CFHR3/CFHR1* deletion (SNP # 10) and rs7517126 (SNP # 11). Interestingly, markers in the third group all showed highly significant associations with plasma CFH concentration, resulting in a large increase in the mean plasma CFH concentration (**Figure 7.10**). Figure 7.10 shows that the change in the direction of effect on AMD susceptibility correlates strongly with the change in the direction of effect on plasma CFH concentration. Three SNPs, rs1048663 (# 3), rs11582939 (# 8) and rs1066420 (# 9) which formed components of the 5-SNP haplotype of Li *et al*. (2006), were found to be effectively neutral in their effects on AMD and plasma CFH concentration (**Figure 7.10**). The results suggest that plasma CFH concentration and *CFH* rs1061170 are acting as combinational factors influencing susceptibility to AMD. Furthermore, plasma CFH concentration was found to explain 1.8% of the AMD disease variance (*P* = 1.7E-03) (**Table 7.10**). *CFH* rs1329428 (# 6) in intron 15 was the only SNP which was found to show a highly significant association with both AMD and plasma CFH concentration (**Figure 7.10**).

Logistic regression analysis of AMD susceptibility factors in the Scottish series was performed using the binary logistic regression methodology implemented in the statistical software package SPSS v.12 and odds ratios (OR) were calculated using the coefficient of the logistic regression (β). The results confirmed *CFH* SNP rs1061170 (Y402H) and *ARMS2* SNP rs10490924 on chromosome 10q26 as the strongest genetic susceptibility loci for AMD (**Figure 7.11**).

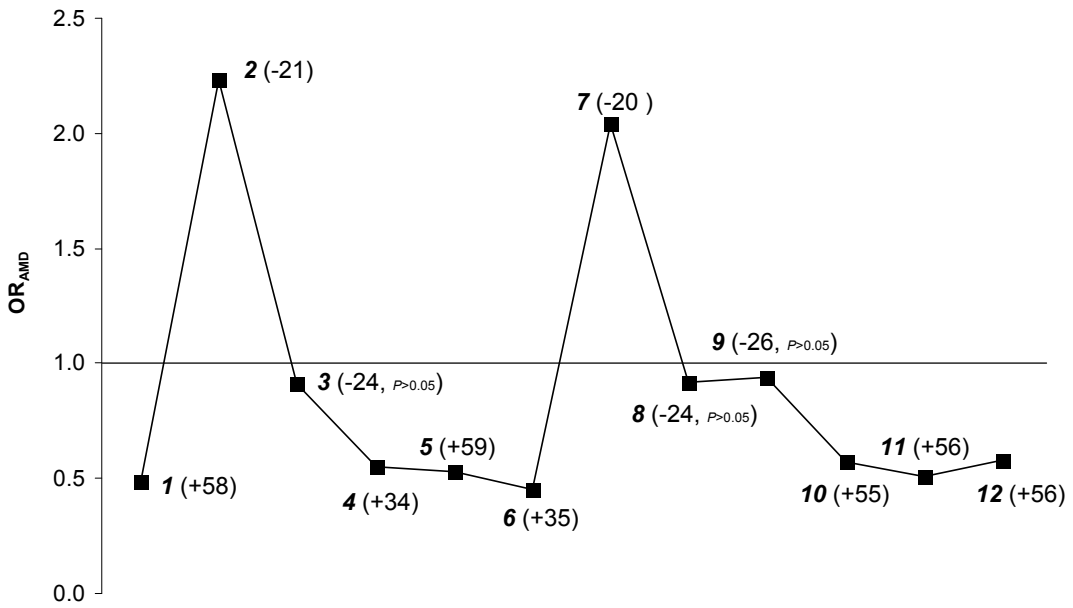


Figure 7.10 The relationship between AMD odds ratios and the effects on plasma CFH concentration associated with single variants.

Diagram shows the sharp change in the direction of effect on AMD susceptibility (odds ratios) correlating with the direction of effect on plasma CFH concentration (shown by β, the numbers in parentheses). The results suggest that plasma CFH concentration and *CFH* rs1061170 (Y402H) could be influencing AMD susceptibility in a combinational manner. Numbers correspond to rs3766404 (***1***), rs1061170 (***2***), rs1048663 (***3***), rs2274700 (***4***), rs6677604 (***5***), rs1329428 (***6***), rs412852 (***7***), rs11582939 (***8***), rs1066420 (***9***), *CFHR3/CFHR1* deletion (***10***), rs7517126 (***11***), rs4086175 (***12***).

Parameter	β	S.E.	t	*P*	η
Plasma CFH concentration	0.0011	0.0004	3.15	1.7E-03	0.018

Dependent variable: AMD status

Table 7.10 Association between plasma CFH and AMD in the Scottish cohort.

Plasma CFH concentration is shown to significantly explain 1.8% of the AMD disease variance in 352 unrelated cases of AMD and 198 controls (*P* = 1.7E-03). The result was obtained using a t-test in a linear regression analysis. β is the regression coefficient of the binary trait and its standard error (SE), η is the proportion of the binary trait variance explained by the covariate (plasma CFH). Analysis was carried out using the software SPSS v.12.

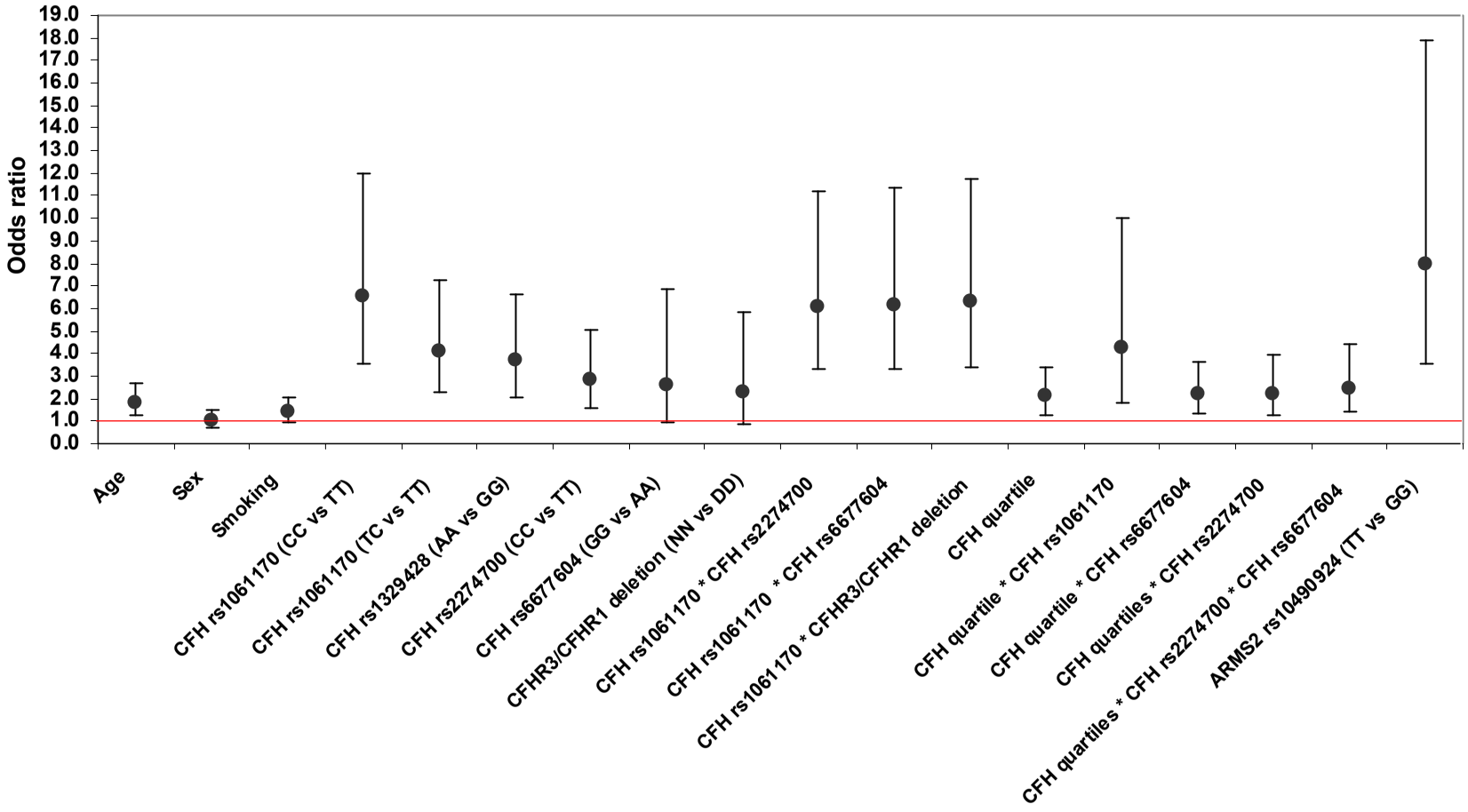


Figure 7.11 Logistic regression analysis of AMD susceptibility factors in the Scottish case-control cohort.

Results of logistic regression analysis in 382 cases of AMD and 201 age-, sex-matched and disease-free controls. Odds ratios for SNPs were calculated based on a reference category, which consisted of the homozygous non-risk genotype. In the case of plasma CFH, the low (0-25%) CFH quartiles were compared with the 75-100th quartile. Error bars represent 95% confidence intervals. Asterisks denote interactions. Analysis was performed using the statistical software SPSS v.12.

The odds ratios (OR) for *CFH* Y402H CC and CT genotypes, relative to the TT genotype, were found to be 6.5 (95% CI 3.5 - 11.9, *P* = 1.5E-09) and 4.1 (95% CI 2.3 - 7.3, *P* = 1.8E-06), respectively. Similarly an OR of 7.9 (95% CI 3.5 - 17.9, *P* = 5.7E-07) was observed for rs10490924 TT homozygotes relative to GG homozygotes. Age and smoking were found to be associated with increased risk of AMD with ORs of 1.8 (95% CI 1.2 - 2.7, *P* = 0.002) and 1.4 (95% CI 0.9 - 2.0, *P* = 0.075), respectively (**Figure 7.11**). Results of the logistic regression analysis of the effects of *CFH* polymorphisms rs1329428, rs2274700 and rs6677604, and the *CFHR3/CFHR1* deletion polymorphism were consistent with the findings of Fisher's exact test of association performed earlier (**Figure 7.11**). Tests of interaction between *CFH* SNPs rs1061170 (Y402H), rs2274700, rs6677604 and the *CFHR3/CFHR1* deletion polymorphism were indicative of an independent effect associated with the rs1061170 (*CFH* Y402H) SNP ($OR_{Y402H*rs2274700}$ = 6.1, 95% CI 3.3 - 11.2, *P* = 7.9E-09; $OR_{Y402H*rs6677604}$ = 6.2, 95% CI 3.3 - 11.4, *P* = 4.9E-09; $OR_{Y402H*deletion}$ = 6.3, 95% CI 3.4 - 11.8, *P* = 7.8E-09) (**Figure 7.11**).

Analysis of the association between plasma CFH and AMD showed that the lowest quartile of plasma CFH concentration (0 - 25%), relative to the highest quartile (75 - 100%), was associated with an increased risk of AMD with an OR of 2.1 (95% CI 1.3 - 3.4, *P* = 0.003) (**Figure 7.11**). Tests of interaction between plasma CFH quartile and rs6677604, rs2274700 and rs6677604*rs2274700 resulted in ORs of 2.2 (95% CI 1.3 - 3.6, *P* = 0.003), 2.2 (95% CI 1.3 - 3.9, *P* = 0.005) and 2.5 (95% CI 1.4 - 4.4, *P* = 0.002). These results suggest that the effects associated with *CFH* rs6677604 and rs2274700 on AMD susceptibility can both be explained by their effects on plasma CFH concentration. Test of interaction between CFH quartiles and *CFH* Y402H showed an OR of 4.3 (95% CI 1.8 - 10.0, *P* = 0.001) compared with CFH quartiles alone (OR = 2.1, 95% CI 1.3 - 3.4) or *CFH* Y402H alone (OR = 6.5, 95% CI 3.5 - 11.9) (**Figure 7.11**). The results suggest an antagonistic model for the effects associated with *CFH* SNP rs1061170 and low plasma CFH concentration since an interaction between the two resulted in a lower odds ratio than the one observed for rs1061170 alone. High CFH concentration (75 - 100% quartile) combined with the *CFH* 1061170 (Y402H) low-risk genotype (TT) resulted in a decreased risk of AMD compared with either alone. Finally, low concentrations of

plasma CFH (0 - 25% quartile) combined with the *CFH* 1061170 (Y402H) low-risk genotype (TT) resulted in an intermediate risk of AMD.

7.2.4.1 Analysis of haplotype association with AMD

Analysis of interactions between *CFH* and flanking SNPs associated with AMD susceptibility was carried out in 382 Scottish cases of AMD and 201 age, sex and ethnicity-matched disease-free controls. Firstly, disease association between 5-SNP haplotypes (rs1048663, rs2274700, rs412852, rs11582939 and rs1066420) reported by Li *et al*. (2006) to be more strongly associated with AMD than *CFH* rs1061170 (Y402H) alone was partially replicated in the Scottish series. The fifth SNP in the haplotype, rs1066420, used by Li *et al*. was excluded from the current analysis due to deviation from Hardy- Weinberg equilibrium ($P_{HWD} < 0.0001$). Haplotype association with AMD was carried out using the haplotype association tool set in PLINK (Purcell *et al*., 2007), which identified one risk and two protective haplotypes (**Table 7.11**). The risk (GCCC, OR = 1.47 P = 6.8E-08) haplotype was observed at the haplotype frequency of 0.516 and the two protective haplotypes, GTTC (OR = 0.654 P = 3.5E-06) and GCTC (OR = 0.205, P = 1.8E-03), were observed at minor haplotype frequencies (MHFs) of 0.294 and 0.016, respectively (**Table 7.11**).

Secondly, conditional haplotype association was used to assess the relationships between AMD susceptibility loci and haplotypes, particular *CFH* SNPs such as rs1061170 (Y402H), rs6677604 and the *CFHR3/CFHR1* deletion polymorphism. A test of AMD association using the 4-SNP haplotypes showed that the effects associated with the GCCC and GTTC haplotypes were not significant when conditioned on the *CFH* Y402H variant ($P_{GCCC} = 0.951$, $P_{GTTC} = 0.200$) (**Table 7.11**). In other words, the effects of the two haplotypes were fully explained by *CFH* rs1061170 (Y402H). The effect associated with the less common protective haplotype, GCTC (MHF = 0.016) was suggestive of being independent from the *CFH* rs1061170 (Y402H) variant (P = 0.004). The effect on AMD susceptibility associated with each of the 4-SNP haplotypes was still significant when conditioned on the *CFHR3/CFHR1* deletion polymorphism (P_{GCCC} = 1.5E-05) and *CFH* rs6677604 (P_{GCCC} = 1.7E-05) indicating that they are not explained by this deletion (**Table 7.11**).

rs1048663	rs2274700	rs412852	rs11582939	Frequency	Odds ratio	P_{HST}	P_{HST} | rs1061170	P_{HST} | deletion	P_{HST} | rs6677604
G	**C**	**C**	**C**	0.516	1.470	6.8E-08	0.951	1.5E-05	1.7E-05
G	T	T	C	0.294	0.654	3.5E-06	0.200	4.5E-04	1.7E-03
G	C	T	C	0.016	0.205	1.8E-03	0.004	5.8E-03	2.8E-03

Table 7.11 Association analysis of 4-SNP haplotypes from Li *et al*. (2006) in Scottish AMD case-control samples.

Four of the five SNPs reported to form AMD susceptibility haplotypes by Li *et al*. (2006) were used in a haplotype association analysis of AMD. Three common haplotype combinations were found to be significantly associated with AMD, one with increased risk, and two with protective effects. Haplotype-specific association tests (HST) were then carried out conditional on *CFH* rs1061170 (Y402H), *CFHR3/CFHR1* deletion and rs6677604. The results show that the 4-SNP haplotypes GCCC and GTTC have no significant effect independent of rs1061170 (*CFH* Y402H). All three haplotypes are independent of the *CFHR3/CFHR1* deletion polymorphism and *CFH* SNP rs6677604. Analysis was carried using the software package PLINK (Purcell *et al*., 2007).

7.2.5 AMD association analysis in the Japanese series

Tests of association between *CFH* and AMD were conducted in a small Japanese case-control series in order to assess whether the polymorphisms affecting plasma CFH in Croatian and Dutch samples were present at a different frequency in the Japanese population, alone explaining the lack of power to detect significant associations between *CFH* and AMD in the Japanese population. DNA samples of 30 unrelated AMD cases (mean age of 77.3 years, 16.7% female) and 58 unrelated age-matched controls (mean age of 76.6 years, 60.3% female) of Japanese origin were genotyped for association between *CFH* or *ARMS2* and AMD. Genotyping was carried out by Dr. T. Iwata at the National Institute of Sensory Organs, in Tokyo due to Japanese restrictions on the export of DNA samples. The AMD grading criteria used for disease diagnosis were described in Section 2.2.

Analysis of SNP frequencies in these unrelated Japanese samples showed that the minor allele frequencies of SNPs rs1061170 (*CFH* Y402H) (minor allele frequency (MAF) = 0.043), rs412852 (MAF = 0.052), rs3766404 (MAF = 0.086) and rs6677604 (MAF = 0.043) were significantly reduced compared to those observed in the Scottish samples (**Table 7.12**). In contrast, higher MAFs were observed for rs10490924 (MAF = 0.336), rs1048663 (MAF = 0.414) and rs11582939 (MAF = 0.422) (**Table 7.12**). A test of association with AMD was carried out using 10 SNPs spanning the *CFH* gene or the *ARMS2* SNP rs10490924 located on chromosome 10q26. Two of the SNPs, rs7517126 and rs4086175, were removed due to low minor allele frequencies (MAF < 0.01). SNP rs1066420 was found to deviate significantly from Hardy-Weinberg equilibrium ($P < 0.001$) and was hence removed. The allelic model was found to be the most suitable model for the test of association after comparing the significance of the allelic, dominant, recessive and genotypic models, using the software package PLINK v1.0 (Purcell *et al*., 2007).

					Allele Frequency Comparisons				
Marker ID	**Minor allele**	**Position**	**Genomic context**	N_{chrom}	**OR (95% C.I.)**	***P***	F_{cases}	$F_{controls}$	$P_{(HWD)}$
rs1061170	**C**	**194925860**	***CFH* (Y402H) exon 9**	**176**	**3.41 (1.06-10.95)**	**0.037**	**0.133**	**0.043**	**0.3797**
rs412852	C	194970330	*CFH* intron 15	176	2.82 (0.93-8.56)	0.077	0.133	0.052	0.4292
rs1066420	G	194987678	4.4 kb downstream of *CFH*	176	1.28 (0.67-2.44)	0.506	0.383	0.328	0.0000
rs1048663	A	194941605	*CFH* intron 9	176	1.08 (0.58-2.03)	0.872	0.433	0.414	0.8294
rs11582939	T	194976780	*CFH* intron 18	176	1.05 (0.56-1.96)	1.000	0.433	0.422	0.6658
rs1329428	A	194969433	*CFH* intron 15	176	0.69 (0.37-1.29)	0.268	0.433	0.526	1.0000
rs3766404	C	194918455	*CFH* intron 6	176	0.56 (0.15-2.11)	0.547	0.050	0.086	0.3797
rs6677604	A	194953541	*CFH* intron 11	176	0.38 (0.04-3.29)	0.665	0.017	0.043	0.0842
rs10490924	**T**	**124204438**	***ARMS2*** (Chrom. 10q26)	**176**	**2.26 (1.19-4.27)**	**0.015**	**0.533**	**0.336**	**0.5068**

Table 7.12 AMD case-control association analysis of unrelated Japanese individuals.

Results of Fisher's exact test of association in 30 unrelated cases of age-related macular degeneration (AMD) and 58 age-matched controls. Seven *CFH* SNPs, including the *CFH rs1061170* (Y402H) variant were used, three of which had shown significant genome-wide association with plasma CFH concentration. The *ARMS2* polymorphism on chromosome 10q26 shown to be associated with AMD in western and Japanese populations (Rivera *et al*., 2005) was also included in the analysis. Genomic positions of the SNPs are based on the March 2006 genome assembly. N, number of chromosomes analysed; OR, odds ratios calculated for the effect of the minor allele; C.I., confidence interval; F, allele frequency in the cases and control; $P_{(HWD)}$, significance value for deviations from Hardy-Weinberg equilibrium. Analysis was carried out using the software package PLINK (Purcell *et al*., 2007).

SNPs rs1061170 (*CFH* Y402H) and rs10490924 (*ARMS2*) were found to be the only significantly associated markers in the Japanese series, using a Fisher's exact test of association after comparing allele frequencies in cases and controls (**Table 7.12**).

The results of the AMD association analysis in the Japanese series showed that *CFH* SNP rs1061170 C allele increased the risk of AMD with an OR of 3.41 (95% CI 1.06 - 10.95, $P = 0.037$). Similarly, rs10490924 (*ARMS2*) T allele was shown to increase the risk with an OR of 2.82 (95% CI 1.19 - 4.27, $P = 0.015$). The $-\log_{10}$ of the significance threshold for the test of association was set at 2.20 ($P = 0.006$) after Bonferroni correction for eight tests. None of the eight markers passed the threshold of significance because of the small sample size. However, the same risk and protective odds ratios as those found in the Scottish AMD series were also observed in the Japanese series. SNPs rs1061170 (*CFH* Y402H) (OR = 3.41, 95% CI 1.06 - 10.95, $P = 0.037$), rs412852 (*CFH*) (OR = 2.82, 95% CI 0.93 - 8.56, $P = 0.077$) and rs10490924 (*ARMS2*) (OR = 2.26, 95% CI 1.19 - 4.27, $P = 0.015$) were found to confer risk. In contrast, SNPs rs1329428 (*CFH*) (OR = 0.69, 95% CI 0.37 - 1.29, $P = 0.268$), rs3766404 (*CFH*) (OR = 0.56, 95% CI 0.15 - 2.11, $P = 0.547$) and rs6677604 (*CFH*) (OR = 0.38, 95% CI 0.04 - 3.29, $P = 0.665$) were found to have protective effects against AMD. The results are therefore similar to those in Caucasian samples although the small sample size reduced the power to detect significant associations with the same SNPs.

7.3 Summary and discussion

In this chapter, the results of a whole-genome association analysis of plasma CFH were presented. Tests of association were carried out using 317,503 biallelic markers as part of the Illumina Hap300 array, in 941 individuals from the island of Vis, 58% of whom were members of 122 families. Polymorphisms not included in the Illumina Hap300 array, which have been reported in the literature to influence susceptibility to AMD, were genotyped separately. The polymorphic deletion of the *CFHR3/CFHR1* genes, reported previously to have a protective effect against AMD (Hughes *et al*., 2006) was genotyped in order to assess the effect of the deletion polymorphism on plasma CFH and its correlation with polymorphisms in the *CFH*

gene. Genotyping was carried out using a real-time quantitative PCR assay which first had to be developed and optimised (**Table 7.1**). TaqMan® SNP genotyping assays and direct sequencing were used for genotyping of the *CFH* rs2274700 and four other SNPs, reported to form the major AMD risk haplotypes (Li *et al*., 2006).

A SNP quality control measure (The Wellcome Trust Case Control Consortium, 2007; Samani *et al*., 2007), was included by careful checking of genotype call rates of all the associated markers using the software Beadstudio (Applied Biosystems), which showed that call rates were all greater than 95%.

Results of genome-wide association analysis of plasma CFH

A *CFH* intronic SNP, rs6677604, was identified as showing the most significant association with plasma CFH concentration in the genome-wide association analysis where it explained 28.3% of the total trait variance (P = 5.19E-56) (**Figure 7.5**). Five other *CFH* SNPs (rs3766404, rs4086175, rs1329428, rs2274700 and rs10801555) showed significant associations with plasma CFH ($P \leq$ 1.6E-07). *CFH* rs2274700, identified by Li *et al*. (2006) to strongly affect AMD susceptibility was the only coding variant showing significant genome-wide association (P = 2.8E-12), explaining about 14% of the adjusted phenotypic variance and encoding an Ala473Ala substitution in exon 10 (**Table 7.4**). The other significant associations occurred outside the *CFH* gene (**Figure 7.5**). These included the polymorphic deletion of the *CFHR3/CFHR1* genes (P = 3.2E-27) and two intergenic SNPs rs7517126 (P = 7.9E-19) and rs4086175 (P = 4.5E-18) (**Table 7.4**). Five other SNPs, including the SNP rs1061170 (*CFH* Y402H) reached suggestive association with plasma CFH ($P \leq$ 1.0E-05) (**Figure 7.5** and **Table 7.4**).

Analysis of the public sequence of the *CFH* locus harbouring the associated markers identified regions with predicted roles in regulation of gene expression as estimated by the regulatory potential (RP) score (**Figure 7.12**). The RP scores were obtained by Dr. P. Gautier (MRC HGU, Bioinformatics Department) from the UCSC Genome Browser (March 2006 assembly) and were estimated based on the alignments of human, chimpanzee, macaque, mouse, rat, dog and cow genomic sequences.

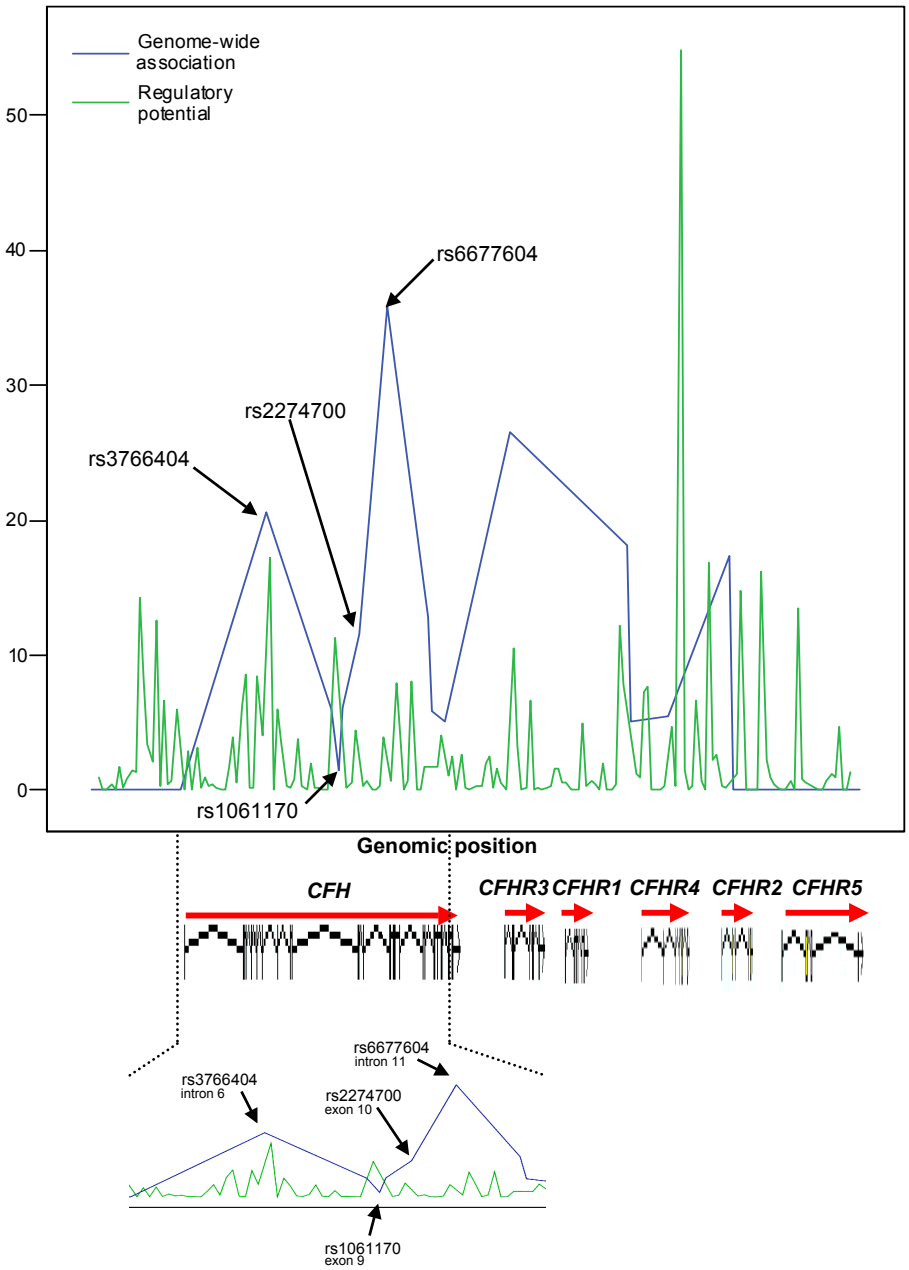


Figure 7.12 Sequence analysis of the 354.5 kb region genetically associated with plasma CFH concentration.

Sequences of DNA with potential regulatory roles in gene expression (regulatory potential (RP)) were identified after analysis of the region showing genome-wide genetic association (blue line) with plasma CFH concentration. Regions of regulatory potential (green line) were detected within 0.5 kb of the association SNPs rs3766404 (150 bp), rs2274700 (100 bp) and rs6677604 (51 bp) as shown in the 60 kb magnified region below. A region of 170 bp with high RP was detected 100 bp downstream of the *CFH* rs1061170 (Y402H) variant. The RP scores were obtained from UCSC (March 2006 assembly) based on the alignments of human, chimpanzee, macaque, mouse, rat, dog and cow sequences. The RP scores were calculated after comparing the frequencies of short alignment patterns between known regulatory elements and neutral DNA. The sensitivity and specificity of RP scores were calibrated on the haemoglobin beta (*HBB*) gene cluster.

The RP scores were calculated after comparing the frequencies of short alignment patterns between known regulatory elements and neutral DNA (Kolbe *et al*., 2004; King *et al*., 2005).

Regions of 50 - 150 bp were identified within 0.5 kb of the most strongly associated *CFH* SNPs, rs3766404, rs2274700 and rs6677604 (**Figure 7.5**) spanning a region of approximately 90 kb (**Figure 7.12**). Regions of high RP score in the vicinity of the highest associated SNPs rs6677604, rs2274700 and rs3766404 implied a possible role in regulation of gene expression associated with these regions (King *et al*., 2005). No correlations between LD break-down and the reduction in significance of association were observed after comparing the LD patterns in the plasma CFH associated locus (**Figure 7.7**) with the association peaks with potential regulatory roles (**Figure 7.12**). Such a correlation, as a result of recombination, would help in narrowing down the region containing a potential regulatory polymorphism. However, the genomic region surrounding the highest associated SNP, rs6677604, is currently being analysed further for a functional role in CFH expression.

Several other markers were also found to be associated with changes in plasma CFH levels ($P \leq$ 1.0E-05), including a coding variant, *F13B* H115R (P = 7.9E-10) and four intronic SNPs in the genes, *ASPM* (1q31), *KCNT2* (1q31), *PTPRN* (2q35), *GBA3* (4p15) and *CAMTA1* (1p36) (**Table 7.3**). However, these associations are likely to occur as a result of the extended LD across the RCA region, possibly resulting from past selection (Voight *et al*., 2006).

Significantly associated SNPs found in the genome-wide association analysis of plasma CFH in the Vis population were replicated in an unselected sample of 500 unrelated Dutch individuals from the Rucphen isolate. *CFH* rs6677604 was again found to be the most significantly associated polymorphism (P = 4.9E-25) explaining 21.8% of the total variance in plasma CFH concentration (**Figure 7.8**, **Table 7.8**). Levels of pairwise LD were generally lower in the Dutch sample set, compared with the Vis (**Figure 7.8**) and Scottish data sets (**Figure 7.9**). The Rucphen community is a recent genetic isolate in the southwest Netherlands for which a comprehensive genealogical database exists (Sayed-Tabatabaei *et al*., 2005). This population was founded in the middle of the 18th century by approximately 150 individuals and was

isolated until the last few decades when rapid growth and minimal immigration took place (Isaacs *et al*., 2007). Since population admixture seems to have been kept to a minimum due to low immigration rates, selection at the *CFH* locus in the Vis and Scottish populations could account for higher levels of pairwise LD in the *CFH* genomic region in these populations (Bamshad & Wooding, 2003). The possibility of selection at the *CFH* locus is analysed further in the final Discussion chapter.

Analysis of pairwise LD between the SNPs associated with plasma CFH concentration, spanning the *CFH* and the *CFH*-related genes, showed that the most strongly associated SNP, rs6677604, was in strong pairwise LD with the *CFHR3/CFHR1* deletion polymorphism (r^2 = 0.84) (**Figure 7.7**). *CFH* rs2274700 (Ala473Ala, exon 10), which has previously been reported to increase susceptibility to AMD, was found to be in high LD with *CFH* rs1329428 intron 15 (r^2 = 0.96). *CFH* SNP rs1061170 (Y402H) (exon 9) showed only borderline genome-wide significance with plasma CFH concentration (**Figure 7.7**).

In order to determine if the effect associated with *CFH* SNP rs6677604 could be explained by a haplotype which accounted for more of the trait variance than the single marker, a multi-marker model was obtained using the 'backward' regression analysis of 27 markers spanning the *CFH* and *CFH*-related genes. A 3-SNP haplotype was obtained, consisting of rs1292487 (located 1.6 kb upstream of *CFH*), rs6677604 (located in *CFH* intron 11) and rs10801575 (located 4.4 kb upstream of *CFHR4*) which together with covariates age and BMI explained 36.9% of the trait variance (**Table 7.5** and **Table 7.6**). However, the 3-SNP haplotype alone explained 29.2% of the plasma CFH variance compared with 28.3% with rs6677604 alone.

Using a haplotype specific test of association, four 3-SNP haplotypes significantly associated with changes in plasma CFH concentration were identified (**Table 7.7**). A common haplotype (MHF = 0.177) containing TAT at sites rs1292487, rs6677604 and rs10801575, resulted in an increase of 55.51 μg/ml in the mean plasma CFH concentration (*P* = 1.1E-26) (**Table 7.7**). This can be compared with the corresponding figure of 59.4 μg/ml for rs6677604 alone (**Table 7.4**). Two other haplotypes, TGT (MHF = 0.105) and TGC (MHF = 0.481) were associated with a decrease of 65.6 μg/ml (*P* = 1.5E-03) and 58.1 μg/ml (*P* = 2.1E-09) in plasma CFH concentrations, respectively (**Table 7.7**).

In order to assess whether the effect associated with the above haplotypes on plasma CFH concentration was independent of the *CFH* rs1061170 (Y402H) SNP and the *CFHR3/CFHR1* deletion polymorphism, a conditional haplotype analysis was carried out. This showed that the effect associated with the TAT haplotype, conditional on rs1061170 (Y402H) was still highly significant (*P* = 2.3E-23) (**Table 7.7**). The most common haplotype, TGC (MHF = 0.481), associated with a decrease of 58.1 μg/ml in plasma CFH concentration was also found to have an independent effect (*P* = 2.3E-03) from that of *CFH* rs1061170 (Y402H) (**Table 7.7**). A test of association between the TAT haplotype and plasma CFH, conditional on the *CFHR3/CFHR1* deletion polymorphism resulted in a reduced but still highly significant *P*-value (*P* = 5.3E-03) indicating that the haplotype explained more of the variation in plasma CFH than the deletion polymorphism (**Table 7.7**), despite a high level of pairwise LD between the SNP rs6677604 and the deletion polymorphism (**Figure 7.7**). The effect on plasma CFH associated with the TGC haplotype was not independent of the deletion, since it was no longer significant (*P* = 0.088) when conditioned on the deletion (**Table 7.7**).

In summary, the TAT haplotype containing the rs6677604 minor allele and associated with increased plasma CFH explained more of the variation in plasma CFH concentration than the *CFHR3/CFHR1* deletion polymorphism (*P* = 5.3E-03). Whether the larger proportion of phenotypic variance explained by the haplotype containing *CFH* rs6677604 is the result of a functional regulatory sequence in the region of rs6677604 can only be determined experimentally.

Relationship between plasma CFH association and AMD

Genetic analysis of 382 Scottish AMD cases and 201 controls, using Fisher's exact test of association, was consistent with the *CFH* rs1061170 (Y402H) variant being the strongest AMD susceptibility locus, with an odds ratio of 2.23 (CI: 1.73 - 2.90, *P* = 7.1E-10) (**Table 7.9**). This variant explained 6.8% of the AMD disease variance in the Scottish series (η^2 = 0.068). Other polymorphisms, including *CFH* SNPs rs3766404 (OR = 0.49, CI: 0.33 - 0.73, *P* = 3.9E-04), rs2274700 (OR = 0.55, CI: 0.42 - 0.73, *P* = 3.3E-05) and rs6677604 (OR = 0.53, CI: 0.38 - 0.75, *P* = 4.6E-04), *CFHR3/CFHR1* deletion polymorphisms (OR = 0.57, CI: 0.41 - 0.81, *P* = 1.4E-03)

and rs7517126 (OR = 0.51, CI: 0.34 - 0.76, *P* = 1.3E-03) were found to have significantly protective effects against AMD (**Table 7.9**, **Figure 7.9**). The proportion of the disease variance explained by *CFH* SNPs rs6677604, rs2274700 and the *CFHR3/CFHR1* deletion polymorphism, individually, was 2.2%, 3.4% and 1.7% respectively (**Table 7.9**).

Table 7.13 summarises the major published work on the genetic association between the *CFH* locus and AMD. Despite six studies which identified the *CFH* rs1061170 (Y402H) variant in exon 9 as the highest genetic susceptibility locus for AMD, with odds ratios ranging between 3.3 - 5.5 (*P* < 1.0E-04) (Haines *et al*., 2005; Edwards *et al*., 2005; Zareparsi *et al*., 2005a; Hageman *et al.*, 2006) three other studies pointed to *CFH* SNP rs2274700 in exon 10 as the most significant single SNP conferring genetic susceptibility to AMD (Li *et al*., 2006 (*P* < 1.0E-30); Hughes *et al*., 2006 (*P* < 1.0E-08); Francis *et al*., 2007 (*P* < 1.0E-03)). Furthermore, haplotypes containing the *CFH* SNP 2274700 (Li *et al*., 2006; Hughes *et al*., 2006) or the *CFHR3/CFHR1* deletion polymorphism and *CFH* SNPs rs6677604 and rs2274700 (Hughes *et al*., 2006) have been shown to influence susceptibility to AMD independently of the *CFH* rs1061170 (Y402H) (**Table 7.13**).

The findings of the present study were consistent with the *CFH* rs1061170 (Y402H) (OR = 2.23, *P* = 7.1E-10) as showing a substantially stronger association with AMD than the *CFH* rs2274700 polymorphism (OR = 0.55, *P* = 3.3E-05) (**Table 7.9**, **Table 7.13**). One study in which *CFH* SNP rs1061170 was found to be more significantly associated with AMD than SNP rs2274700 was conducted using two independent series of unrelated American individuals of European descent (Hageman *et al*., 2005) (**Table 7.13**). The lower significance of association between *CFH* rs2274700 and AMD in the above American and Scottish series could be explained by different selective forces acting in these populations, but the Irish and the North American populations of Hughes *et al*. (2006) and Li *et al*. (2006) have essentially identical population histories, so this seems unlikely. Selection at the *CFH* locus has been implicated as a major cause of diversity in the *CFH* genomic region (Hageman *et al*., 2006).

Reference	Highest associated SNP (Genomic location)	Minor allele (MAF aff., unaff)	*CFH* rs1061170 OR (CI, *P*) M.A. (MAF aff., unaff.)	*CFH* rs2274700 OR (CI, *P*) M.A. (MAF aff., unaff.)	*CFH* rs6677604 OR (CI, *P*) M.A. (MAF aff., unaff.)	*CFHR3/CFHR1* deletion OR (CI, *P*) M.A. (MAF aff., unaff.)	*N* (aff., unaff.)
Scottish AMD (present study)	*CFH* rs1061170 (Exon 9)	C (0.54, 0.34)	2.2$^{\zeta}$ (1.7-2.9, 7.1E-10)	0.55 (0.4-0.7, 3.3E-05) T (0.25, 0.37)	0.53 (0.3-0.7, 4.6E-04) A (0.11, 0.19)	0.57 (0.4-0.8, 1.3E-03) D (0.12, 0.20)	382, 201
Haines *et al*., 2005	*CFH* rs1061170 (Exon 9)	C (0.94, 0.46)	3.3$^{\psi}$ (1.8-6.2, P<0.0001)	n.a.	n.a.	n.a.	24, 24
Edwards *et al*., 2005	*CFH* rs1061170 (Exon 9)	C (0.55, 0.34)	2.7$^{\psi}$ (1.9-3.9, 3.6E-08)	n.a.	n.a.	n.a.	224, 134
Klein *et al*., 2005	*CFH* rs380390 (Exon 15)	C (0.30, HapMap)	n.a.	n.a.	n.a.	n.a.	96, 50
Zareparsi *et al*., 2005	*CFH* rs1061170 (Exon 9)	C (0.61, 0.34)	5.5$^{\psi}$ (3.5-8.6, P<10^{-24})	n.a.	n.a.	n.a.	616, 275
Hageman *et al*., 2006 (Iowa)	*CFH* rs1061170 (Exon 9)	C (0.59, 0.34)	2.8$^{\zeta}$ (2.1-3.8, 2.1E-12)	3.4 (2.3-5.1, 3.1E-09) T (0.20, 0.46)	n.a.	n.a.	404, 131
Hageman *et al*., 2006 (Columbia)	*CFH* rs1061170 (Exon 9)	C (0.54, 0.34)	2.2$^{\zeta}$ (1.8-2.7, 1.6E-13)	2.1 (1.7-2.6, 1.6E-11) T (0.29, 0.46)	n.a.	n.a.	550, 275
Li *et al*., 2006	*CFH* rs2274700 (Exon 10)	T (0.16, 0.45)	LRT 110 (P<10^{-25}) C (0.62, 0.34)	LRT 135.4$^{\zeta}$ (P<10-30)	LRT 31.2 (P<10^{-07}) A (0.10, 0.20)	n.a.	544, 268
Hughes *et al*., 2006*	*CFH* rs2274700 (Exon 10)	n.a.	χ^2 20.5 (*P*=5.9E-06)	χ^2 35.2z (P=2.9E-09)	χ^2 19.7 (*P*=8.9E-06)	0.35 (*P*=1.0E-05)** (7.8, 19.4)	173, 170
Francis *et al*., 2007	*CFH* rs2274700 (Exon 10)	n.a.	2.4 (2.1-2.7, 1.0E-04) C (0.59, 0.36)	3.3$^{\zeta}$ (2.7-3.9, 3.0E-04)	n.a.	n.a.	211, 183

Table 7.13 Overview of the data published on genetic association analysis of the *CFH* locus.
Key: ζ, allele frequency comparison; ψ, logistic regression referenced on the homozygous non-risk genotype; *, this data was also available from 19 cases of AMD and 40 controls from the Irish population; **, in the context of a protective *CFH* haplotype; OR, odds ratio; CI, 95% confidence interval; M.A., minor allele; MAF, minor allele frequency; LRT, likelihood ratio test; n.a., not available; *N*, number of affected (aff.) and unaffected (unaff.) individuals. Studies which have carried out *CFH* haplotype association with AMD are highlighted in bold.

Secondly, the high association of the *CFH* rs2274700 in the US population reported by Li *et al*. (2006) may be the result of population substructure leading to spurious associations (Wellcome Trust Case Control Consortium, 2007). This hypothesis is unlikely since three independent studies identified the *CFH* rs2274700 polymorphism as being more strongly associated with risk of AMD than *CFH* rs1061170 (**Table 7.13**).

Markers associated with AMD after Bonferroni correction for multiple testing were representative of three LD groups ($r^2 > 0.50$) in the Scottish AMD series (**Figure 7.9**). These LD groups contain SNPs rs1061170 (Y402H) and rs412852 (LD group 1), which showed borderline association with plasma CFH concentrations, rs1329428 and rs2274700 (LD group 2), and rs3766404, rs6677604 and *CFHR3/CFHR1* deletion polymorphism (LD group 3), which showed highly significant associations with plasma CFH concentrations (**Figure 7.9**). It was therefore hypothesised that the polymorphisms in LD groups 2 and 3 affect AMD susceptibility independently of the *CFH* rs1061170 (Y402H) by significantly altering plasma CFH concentrations.

To test the above hypothesis, conditional haplotype analyses were carried out, testing for non-independence of marker associations to disease depending on their presence in different LD groups. The effect of a 5-SNP haplotype, containing rs1048663, rs2274700, rs412852, rs1066420 and rs11582939 reported by Li *et al*. (2006) on the AMD association in the Scottish case-control sample, conditional on other SNPs such as rs1061170 (*CFH* Y402H), rs6677604 or the *CFHR3/CFHR1* deletion polymorphism was analysed. The results showed that the effects associated with the GCCC and GTTC haplotypes were not significant when conditioned on the *CFH* Y402H variant, consistent with non-independence ($P_{GCCC} = 0.951$, $P_{GTTC} = 0.200$) (**Table 7.11**). The effect associated with the less common haplotype, GCTC (MHF = 0.016), was consistent with partial independence from the *CFH* Y402H variant ($P = 0.004$). A previous study has also shown that the effect of the GCCC haplotype is no longer significant when conditioned on *CFH* rs1061170 (Y402H) ($P_{haplotype} \mid$ *CFH* Y402H = 0.938) (Li *et al*., 2006). However, in the same study, the GTTC haplotype is shown to confer susceptibility independently of the *CFH* Y402H, a finding which is inconsistent with the present study. Furthermore, Li *et al*. showed

that the effect associated with the protective haplotype, GCTC, was non-significant when conditioned on the *CFH* Y402H variant ($P_{haplotype}$ | *CFH* Y402H = 0.396) which was again inconsistent with this study.

The results of the conditional haplotype analysis in the Scottish AMD series also showed that the effects of the 4-SNP haplotypes were significant when conditioned on both the deletion of *CFHR3/CFHR1* genes (P_{GCCC} = 1.5E-05) and the *CFH* SNP rs6677604 (P_{GCCC} = 2.8E-02) (**Table 7.11**) and were therefore independent of these polymorphisms.

In summary, the results suggest that *CFH* rs2274700, rs6677604 and the *CFHR3/CFHR1* deletion polymorphism are strongly associated with plasma CFH concentrations as a result of high pairwise linkage disequilibrium, and collectively influence AMD susceptibility ($P < 0.05$) independently of the SNP rs1061170 (*CFH* Y402H).

Genetic analysis of AMD was carried out in a Japanese series aiming to assess whether the polymorphisms affecting differences in plasma CFH concentration in the Croatian and Dutch populations were present at different frequencies in the Japanese population, perhaps explaining the lack of power in detecting significant associations between *CFH* and AMD in the Japanese population. The analysis was carried out using a Fisher's exact test of association in a small series of 30 unrelated cases of AMD and 58 unrelated age-matched controls. The analysis also included the *ARMS2* rs10490924 polymorphism on 10q26, previously found to affect susceptibility to wet AMD in the Japanese population (Yoshida *et al*., 2007; Kondo *et al*., 2007). *CFH* rs1061170 (Y402H) was found to confer the largest risk of AMD (OR = 3.41, 95% CI 1.06 - 10.95, P = 0.037) followed by the *ARMS2* SNP rs10490924 (OR = 2.26, 95% CI 1.19 - 4.27, P = 0.015) (**Table 7.12**). Other *CFH* markers did not show significant associations with AMD, probably due to the small number of samples analysed and lack of power.

The allele frequency of the *CFH* Y402H variant has been found to show ethnic variation, with lower frequencies in Chinese and Japanese populations, compared to the Caucasians (Chen *et al*., 2006; Uka *et al*., 2006; Lau *et al*., 2006; Okamoto *et al*., 2006). AMD association studies of the 1q31 and 10q26 loci have revealed conflicting results. Most studies have failed to detect significant association

between *CFH* Y402H and wet AMD in the Chinese (Chen *et al*., 2006) or Japanese (Uka *et al*., 2006; Gotoh *et al*., 2006; Okamoto *et al*., 2006) populations. Absence of a *CFH* association has also been reported for dry AMD in the Japanese population (Fuse *et al*., 2006). However, a significant association between wet AMD and *CFH* Y402H has been published for the Chinese population (Lau *et al*., 2006). On the other hand, variants in the 10q26 locus have been shown to significantly confer susceptibility to wet AMD in both Chinese (Dewan *et al*., 2006) and Japanese (Yoshida *et al*., 2007; Kondo *et al*., 2007) populations.

Considering the large proportion of the AMD cases in the present study had presented with the 'wet' form of the disease (23 out of 30 cases, **Table 2.1**), the results are consistent with significant associations between *CFH* Y402H and *ARMS2* rs10490924 variants and 'wet' AMD. Comparison of the allele frequencies between the control populations of the Japanese and Scottish samples showed significant reduction in the minor allele frequencies of the *CFH* SNPs rs1061170 (Y402H), rs6677604, rs412852 and *ARMS2* rs10490924 in the Japanese population compared with its Scottish counterpart (**Table 7.14**). Despite the reduction in the minor allele frequency of the *ARMS2* SNP rs10490924 in the Scottish population compared with the Japanese, this SNP showed the strongest association with AMD in the Scottish population (OR = 2.6, P = 9.0E-11) (**Table 7.9**). Despite a significant reduction in the frequency of the *CFH* Y402H risk allele C in the Japanese population, this variant showed the strongest association with AMD in this population (OR = 3.4, P = 0.037). The minor allele frequencies of three SNPs from the 5-SNP haplotype identified by Li *et al*. (2006) increased in the Japanese control population compared with the Scottish sample (**Table 7.14**). The findings of the present study are consistent with a significant association between *CFH* rs1061170 (Y402H) and 'wet' AMD in the Japanese population. The lack of a sufficiently large sample size was the limiting factor in detecting significant associations in the available Japanese AMD series.

In conclusion, genome-wide association analysis of plasma CFH resulted in significant associations with several polymorphisms close to and within the *CFH* gene. The strongest association was with an intronic SNP within *CFH*, which explained 28% of the total trait variance (P < 1.0E-50) (**Figure 7.4**, **Table 7.3**).

	rs1061170	rs3766404	rs6677604	rs412852	rs1066420	rs1048663	rs11582939	rs10490924
Minor allele	C	C	A	C	G	A	T	T
MAF (Japanese controls)	0.043	0.086	0.043	0.052	0.328	0.414	0.422	0.336
MAF (Scottish controls)	0.345	0.152	0.194	0.406	0.154	0.177	0.177	0.191
P-value$_{2\text{-tailed}}$	**2.4E-12**	0.067	**2.2E-05**	**4.8E-15**	**9.0E-05**	**3.1E-07**	**1.6E-07**	**0.0025**

Table 7.14 Comparison of the minor allele frequencies between the Scottish and Japanese control populations.

The minor allele frequencies of some of the SNPs which showed associations with AMD in this study were compared between the Japanese (116 chromosomes) and the Scottish (402 chromosomes) control populations as a measure of the minor allele frequencies (MAF) of these SNPs in the above populations. Results of the Fisher's exact test of allele comparisons showed highly significant differences in allele frequencies between the two populations except for *CFH* SNP rs3766404 ($P < 0.05$). The main limitation in this comparison is that the Scottish sample was disease(AMD)-free which may produce a bias in allele frequencies. Significant *P*-values are highlighted in bold.

The association was also replicated in a Dutch sample set (**Figure 7.8**). The 3-SNP TAT haplotype identified by 'backward' regression was associated with a 55.5 μg/ml increase in plasma CFH concentration (**Table 7.7**) compared with a 59.4 μg/ml increase associated with the single SNP rs6677604 (**Table 7.4**) showing that the haplotype did not explain more of the effect on plasma CFH than the single SNP alone. Conditional haplotype analysis showed that the effect of this (TAT) haplotype on CFH concentration was independent of the *CFH* Y402H variant, and of the *CFHR3*/*CFHR1* deletion polymorphism (**Table 7.7**), previously shown to affect AMD susceptibility (Hughes *et al*., 2006).

Genetic analysis of 382 Scottish AMD cases and 201 disease-free controls was consistent with *CFH* SNP rs1061170 (Y402H) being the strongest AMD susceptibility locus, which explained 6.8% of the disease variance (**Table 7.9**). Variation in plasma CFH concentration was found to explain up to 1.8% of the variation in susceptibility to AMD (**Table 7.10**) with an odds ratio of 2.1 (95% C.I. 1.3 - 3.4, $P = 0.003$) (**Figure 7.11**). The results indicate that SNPs, which are strongly associated with plasma CFH concentration collectively influence AMD susceptibility ($P < 0.05$) independently of the *CFH* rs1061170 (Y402H) polymorphism.

Although the 3-SNP haplotype did not significantly narrow down the association region, and it did not account for a significantly larger effect on plasma CFH than the single SNP (rs6677604), it is still plausible that the combination of the three SNPs, or variants associated with them, affect plasma CFH by regulating gene expression. Analysis of the regulatory potential of the sequence within the *CFH* locus also identified many non-coding regions with potential roles in regulating gene expression.

In the final Discussion (Chapter 9) a model is proposed for AMD susceptibility based on the findings of this study, which aims to dissect the complex architecture of the disease, in particular in the context of plasma CFH.

CHAPTER 8

ANALYSIS OF PROTEINS BY MASS SPECTROMETRY

8.1 Introduction

Development of techniques capable of large-scale protein quantitation would be greatly advantageous in the analysis of quantitative traits that are intermediate phenotypes for human diseases. The goal is the large-scale quantitation and screening of proteins which indicate differences in gene expression or protein turnover in the general population, using proteins of known relevance to human disease.

A number of techniques have been employed for the analysis of simple mixtures of proteins and peptides. These include photometric, radiometric, immunological and electrophoretic techniques. However, large-scale detection and analysis of proteins has only become possible due to the advent of mass spectrometric techniques. Mass spectrometry (MS) methods are commonly based on sample ionisation and fragmentation followed by detection. The main features of some of the classical techniques, such as MALDI-TOF, capillary electrophoresis and HPLC are their abilities to confidently detect and characterise protein samples, based on digestion, ionisation or fragmentation products. More sophisticated methodology such as liquid chromatography coupled with tandem MS/MS facilities, and Fourier transform ion cyclotron resonance (FT-ICR) have provided significantly higher resolution for determining protein masses with sub-femtomole sensitivity. The ability of some of these techniques, in particular capillary electrophoresis coupled with FT-ICR MS/MS, in relative-quantitation of proteins in samples such as urine has been assessed and is promising (Weissinger *et al*., 2004). The most widely accepted procedure for the relative quantitation of proteins in human tissue is iTRAQ™ coupled with MS/MS. (Ross *et al*., 2004; DeSouza *et al*., 2005). Relative quantitaion using iTRAQ™ is performed via tandem mass spectrometry (MS/MS) after labelling with four possible tags placed on primary amines. Quantitation is performed via the differences in abundances of four product ions with mass/charge ratios of 114, 115, 116 and 117 Da, that are each cleaved from one of the four possible tags.

Despite the large amount of information relating to disease in human plasma, quantitative analysis of the plasma proteome has faced great challenges. One of the major problems which needs to be overcome prior to any type of large-scale analysis

of plasma is its complexity. Plasma contains low abundance proteins, which are mainly products of tissue leakage at pg/ml levels, in the presence of several highly abundant proteins. Furthermore, the high abundance proteins often mask the lower abundance ones, making their detection more difficult.

The aim of this chapter is to assess the ability of the various mass spectrometry techniques to detect and quantitate plasma proteins. Both methods of ionisation, namely MALDI-TOF MS and electrospray ionisation (ESI) FT-ICR, were assessed. Capillary electrophoresis was also used as a method of fractionation prior to ionisation and MS analysis. Finally, measurement of plasma CFH concentrations by ELISA were used to assess the ability of ion exchange chromatography, coupled with nano-scale liquid chromatography (nLC) joined with tandem MS (MS/MS) to determine relative quantities of CFH protein in different plasma samples.

Several 'upstream' methodologies were assessed in their efficiency and ability to overcome the problem of complexity and "interference" by high abundance proteins in plasma. Interference refers to saturation effects as highly abundant proteins suppress the detection of the low abundance protein fragments. Methods to overcome the problem of interference included immunodepletion of high abundance proteins, chromatography, ultrafiltration and hydrophobic interaction chromatography to selectively collect proteins and peptides of low abundance.

8.2 Results

8.2.1 Immunodepletion of high abundance plasma proteins

Immunodepletion was carried out using a multiple affinity removal spin column (cartridge) as described in Section 2.4.8.1. The column is designed to remove greater than 98-99% of six interfering high-abundance proteins, including albumin, IgG, IgA, transferrin, haptoglobin and α_1-antitrypsin from the human plasma. Ten microlitres of crude plasma was used as the starting material and each sample was co-eluted twice, together with 1-2 µl of α-cyano-4-hydroxycinnamic acid (α-cyano) matrix onto a MALDI plate (**Figure 8.1**). The spectrum of crude human plasma was obtained by MALDI-TOF MS capturing the mass range of 3.5 kDa to 100 kDa. The spectrum was mainly dominated by singly and doubly charged ions of serum albumin with a mass to charge ratio (m/z) of 66.5 kDa and 33.2 kDa respectively.

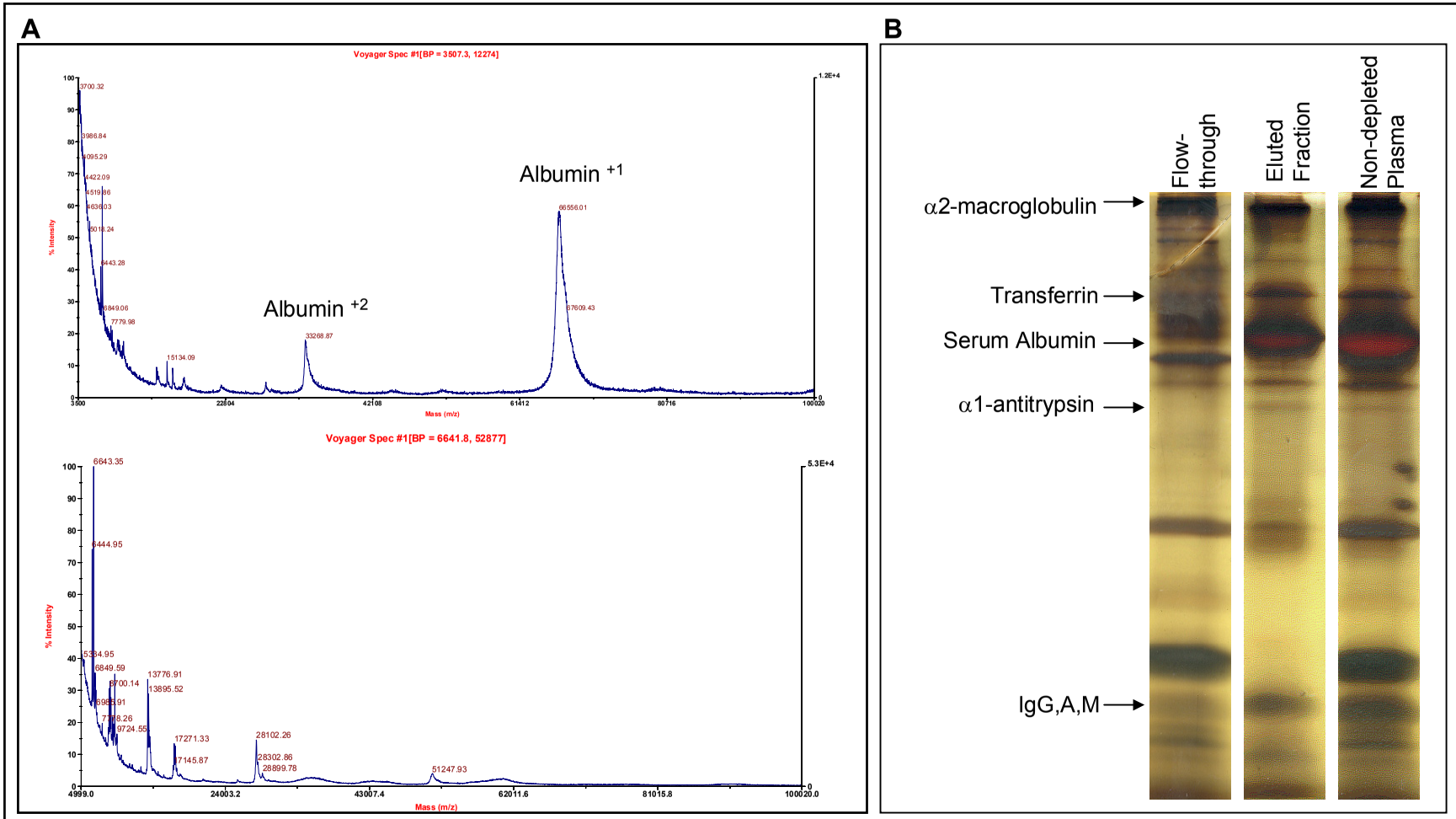


Figure 8.1 Mass spectra of crude and immunodepleted human plasma.

Immunodepletion of high abundance plasma proteins confirmed by MALDI-TOF MS (A) and silver staining (B). **A.** Mass spectra obtained by MALDI-TOF MS, after immunodepletion (bottom) of a sample of crude human plasma (top). **B.** The silver stained plasma proteome separated by SDS-PAGE shows the absence of the depleted proteins in the flow-through and their presence in the eluted fraction and crude plasma. Spectra were obtained using a Voyager-DE™ STR instrument (Applied Biosystems).

After immunodepletion of the most abundant proteins in plasma, using the depletion column, the absence of albumin, in both its ionic forms, was confirmed. This improved the resolution for detecting other proteins present in the smaller mass range of 5 to 25 kDa.

However, as the mass spectra of immunodepleted plasma obtained by MALDI-TOF MS show (**Figure 8.1-A),** immunodepletion alone does not result in detection of many proteins, especially in the mass range 30 kDa to 100 kDa which is believed to be the favourable mass range for protein ionisation using MALDI-TOF MS.

The ability of the multiple affinity column in eliminating the most abundant plasma proteins was also assessed by silver staining SDS polyacrylamide gels (**Figure 8.1**). The silver stain, with detection limits of 0.1-1.0 ng of protein confirmed the absence of the high abundance proteins, after separation by SDS-PAGE according to their molecular weight. Comparison of the flow-through to the non-depleted plasma showed the absence or significant reduction in the intensity of alpha-2-macroglobulin, transferrin, serum albumin, alpha-1 antitrypsin and immunoglobulin G, A and M bands. The flow-through fraction contained the remaining 10% of plasma proteome consisting of the low abundance fraction. The presence of the same bands at higher intensity in the eluted column fraction following immunodepletion illustrated the efficiency of the depletion system.

After assessing the efficiency of the multiple affinity column in depleting the high abundance fraction of plasma, a number of techniques were employed to further optimise detection and to assess the quantitative aspect of mass spectrometry.

8.2.2 MALDI-TOF mass spectrometry

For the analysis of plasma using MALDI-TOF MS, the mass range was divided into three parts, the small mass range (1 kDa < m/z <5 kDa), the medium mass range (5 kDa < m/z < 100 kDa) and the large mass range (100 kDa < m/z < 200 kDa). This was due to the preferential ionisation capacity of MALDI, with the medium mass range having the ideal molecular mass for ionisation. The aim was to assess the use of plasma CFH and Aβ levels measured previously by ELISA as internal calibrators in the medium and small mass ranges, respectively. If successful, the relative

abundance of other peaks could be deduced based on the intensities of the observed CFH and Aβ peaks. Each MALDI-TOF MS experiment also included three negative controls which consisted of the spotted matrix alone.

Since immunodepletion alone did not improve the m/z resolution obtained from MALDI-TOF MS, different techniques were optimised and applied to the three assigned mass ranges, and the results were as follows.

8.2.2.1 MALDI-TOF MS of the small mass range

Crude plasma was diluted, acidified and concentrated using 10 µl ZipTip™ pipette tips, according to the manufacturer's instructions (Millipore). After fractionation by chromatography resins embedded in the ZipTip™$_{\mu\text{-C18}}$ tips, the sample was co-eluted with α-cyano-4-hydroxycinnamic acid (α-cyano) matrix onto the MALDI plate and subject to mass spectrometric detection, as described in Methods. Each plasma sample was analysed independently three times on the same plate.

This fractionation procedure resulted in improved spectra with higher signal to noise ratio (**Figure 8.2**). Mass spectra from plasma treated in this way revealed a larger number of peaks compared with spectra from crude or immunodepleted samples. The highest peak was detected at m/z 2,753.4 and was present in sample replicates. However, neither the intensity nor the area under the peak, were found to be reproducible between sample replicates. Furthermore, due to the limitation of MALDI-TOF with regard to protein identification, the identity of this peak could only be postulated based on its m/z ratio. The peak was believed to be a fragment of human immunoglobulin in plasma, using the Mascot protein search engine. Other reproducible peaks were detected at m/z 1,896, 2,937 and 3,473. No peaks in the region of 4.3 to 4.5 kDa were detected, indicating that the low concentrations of amyloid-β species at the pg/ml level were preventing them from being detected by this technique.

Ultrafiltration of plasma using 10 kDa molecular cut-off filters (Millipore), which prevented passage of molecules larger than 10 kDa, followed by acidifying samples with 1% TFA prior to desalting and concentrating with µC18 ZipTip™ pipette tips, resulted in an even larger number of peaks in the small protein size window (**Figure 8.3**).

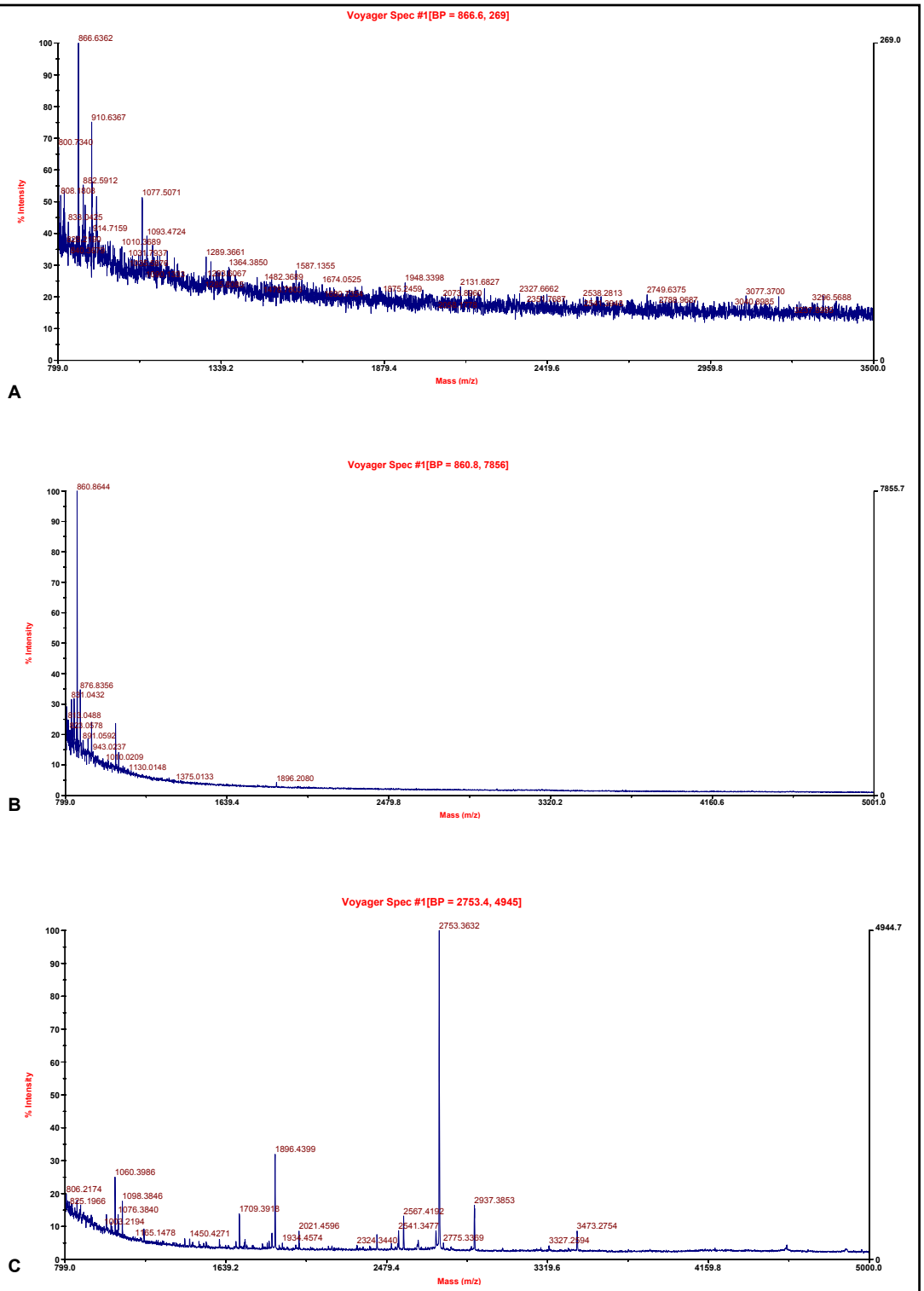


Figure 8.2 MALDI-TOF MS spectra of the small mass range showing results of ZipTip™ treatment of plasma.

Mass spectra of the treated human plasma showing an improvement in the mass spectrometry signal associated with ZipTip™ treatment compared with immunodepletion. Mass spectra were obtained for the small mass range of 1 kDa to 5 kDa using MALDI-TOF MS. **A.** The spectrum of crude human plasma shows high levels of background noise. **B.** Immunodepletion of high abundance proteins results in a better resolution. **C.** The highest signal to noise ratio is observed for the diluted, concentrated and ZipTip™ purified samples. Spectra were obtained using the MALDI-TOF MS Voyager-DE™ STR (Applied Biosystems) in a linear mode.

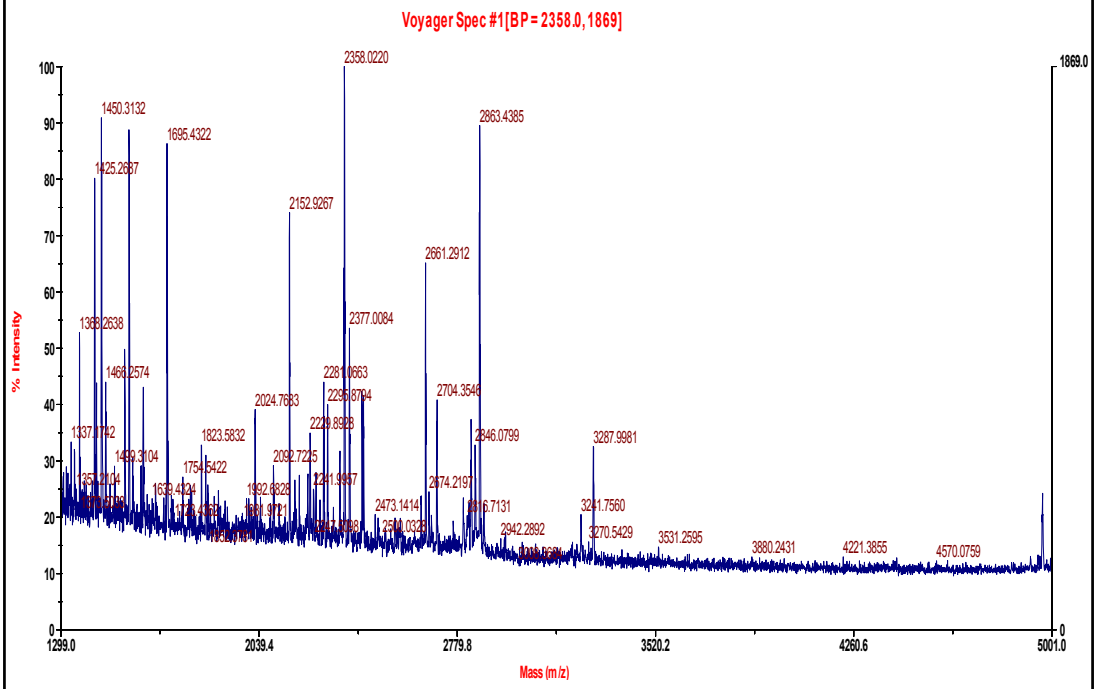


Figure 8.3 MALDI-TOF MS spectrum of human plasma after ultrafiltration using 10K MW cut-off columns.

Mass spectrum of human plasma after ultrafiltration followed by ZipTip™ treatment of crude plasma using 10K MW cut-off columns showing improvements in the number of detectable peaks and the intensity of the m/z signal compared with spectra of immunodepleted human plasma. Mass spectra were obtained for the small mass range of 1.3 kDa to 5 kDa using MALDI-TOF MS. Ultrafiltration of 2 ml of crude plasma was followed by diluting, acidifying and concentrating with ZipTip™. Spectra were obtained using the MALDI-TOF MS Voyager-DE™ STR (Applied Biosystems) in a linear mode.

As before, samples were spotted onto a MALDI plate in triplicate, using 1-2 μl of α-cyano matrix. This technique produced more promising results, with a larger number of detectable peaks. However, large variation was still observed between replicates of the same sample. The high level of variation between replicates could be the result of the number steps involved in the purification of plasma samples introducing high levels of inter-experimental variation.

The ability of MALDI-TOF MS to quantitate proteins and peptides in the small mass range was assessed by spiking plasma samples with known concentrations of synthetic human Aβ42 (Sigma). With human Aβ42 plasma concentrations ranging from 10 pg/ml to 200 pg/ml, lack of detection of a peak at 4.5 kDa had previously indicated the inability of MALDI to detect Aβ42 at physiological levels. Firstly, single polypeptide analysis was carried out using Aβ42 at concentrations of 200 pg/ml, 500 pg/ml, 1000pg/ml, and 1000 ng/ml diluted in HPLC dH_2O. After treatment with C18 ZipTip™, 1 μl of purified peptide was co-eluted with 1 μl of α-cyano matrix and spotted on a MALDI plate for MALDI-TOF MS analysis. The low sensitivity and resolution of this techniques resulted in the lack of detection of the Aβ42 at concentrations of 200 pg/ml, 500 pg/ml and 1000 pg/ml (**Figure 8.4**). Detection of the peptide at 1000 ng/ml, considerably higher than its normal physiological or even disease-related concentrations, was achieved successfully and with high confidence (signal intensity, 4.2E+04). **Figure 8.4-D** also showed that in the presence of the peptide at 1000 ng/ml, peaks which had dominated the spectrum at lower concentrations of Aβ42 (**Figure 8.4 A-C**), became part of the background noise. This was frequently observed in MALDI-TOF analysis and was another indication of the dominating effect of proteins of high concentration relative to other proteins in the same mass range.

In order to assess the reproducibility of Aβ42 peptide detection, peptides diluted in HPLC dH_2O at concentrations of 200 ng/ml, 500 ng/ml and 1000 ng/ml were spotted onto a MALDI plate as before and analysed by MS. Each sample was analysed independently in 5 replicates in the presence of negative controls consisting of the matrix alone. The results showed high levels of variation between replicates. Furthermore, no direct relationship between the starting concentration of Aβ42 peptide and the peak height was observed (**Figure 8.5**).

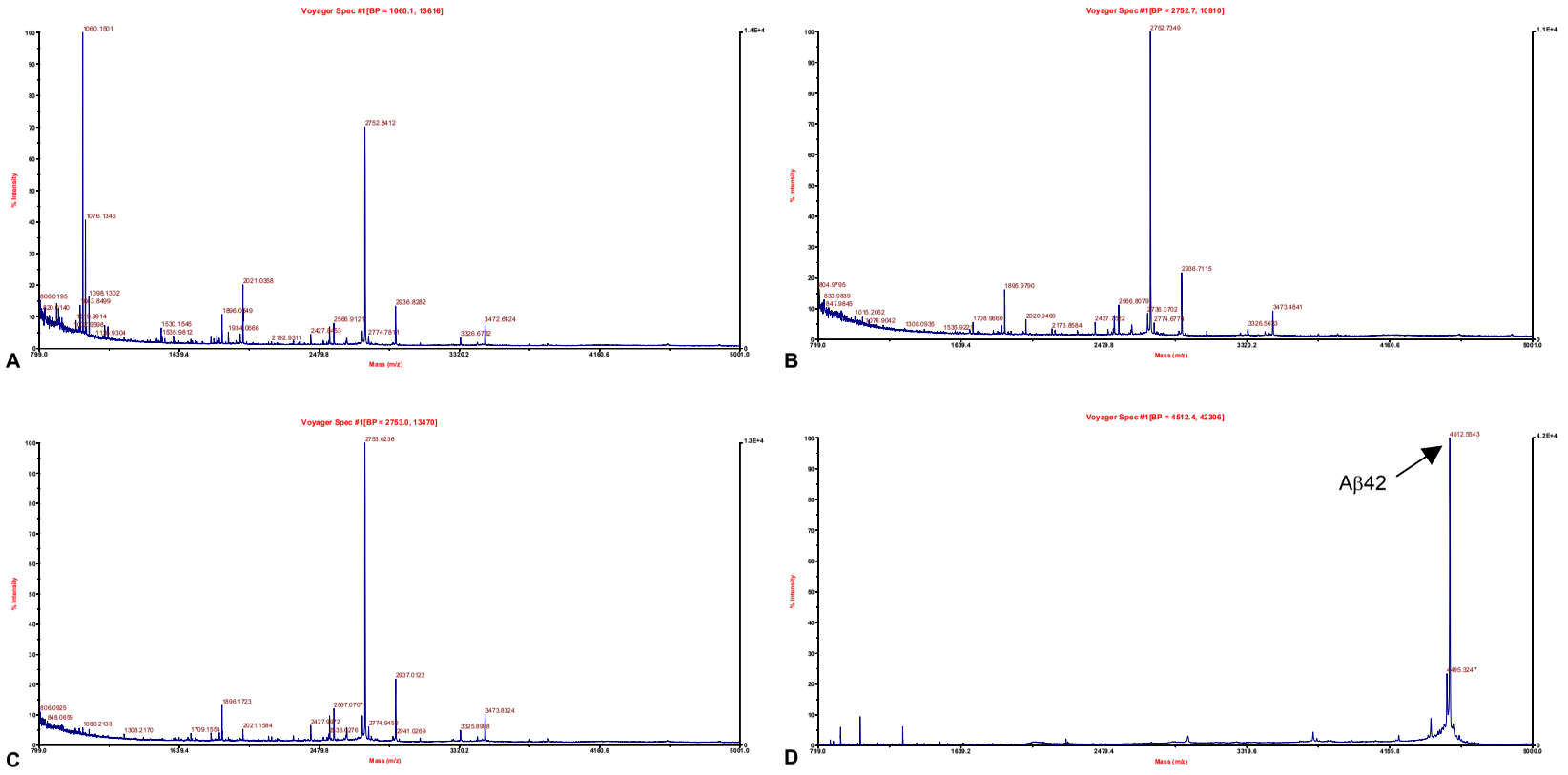


Figure 8.4 Assessment of MALDI-TOF MS in quantitative determination of Aβ42 peptide.

Spectra of human plasma spiked with varying concentrations of human Aβ42 peptide showing lack of detection of this peptide by MALDI-TOF MS at near-physiological concentrations (10-200 pg/ml). Aβ42 peptide diluted in HPLC dH_2O at concentrations of 200 pg/ml **(A)**, 500 pg/ml **(B)**, 1000 pg/ml **(C)** and 1000 ng/ml **(D)** was treated with ZipTip™ as before. 1 μl of the concentrated peptide was co-eluted with alpha-cyano matrix and analysed by MALDI-TOF MS using a Voyager-DE™ STR (Applied Biosystems) in a linear mode.

Aβ42 peptide at a concentration of 200 ng/ml did not result in improved resolution (data not shown). Increasing the concentration to 500 ng/ml resulted in a slight increase in the signal at m/z of 4,513 but with low certainty (5,528.7) (**Figure 8.5**). A further increase in the concentration of the eluted peptide to 1000 ng/ml resulted in a major improvement in the signal at m/z of 4,513 (3.0E+04). **Figure 8.5** also showed a feature of MALDI-TOF which is detrimental to quantitative detection of protein peaks. This is that the intensities of other peaks in the region, for example the peak at m/z of 2,752 (**Figure 8.5-A**) and those at m/z of 931, 1,046 and 1,296 (**Figure 8.5-B**) were greatly influenced by the presence or absence of the Aβ42 peptide peak (m/z of 4,513).

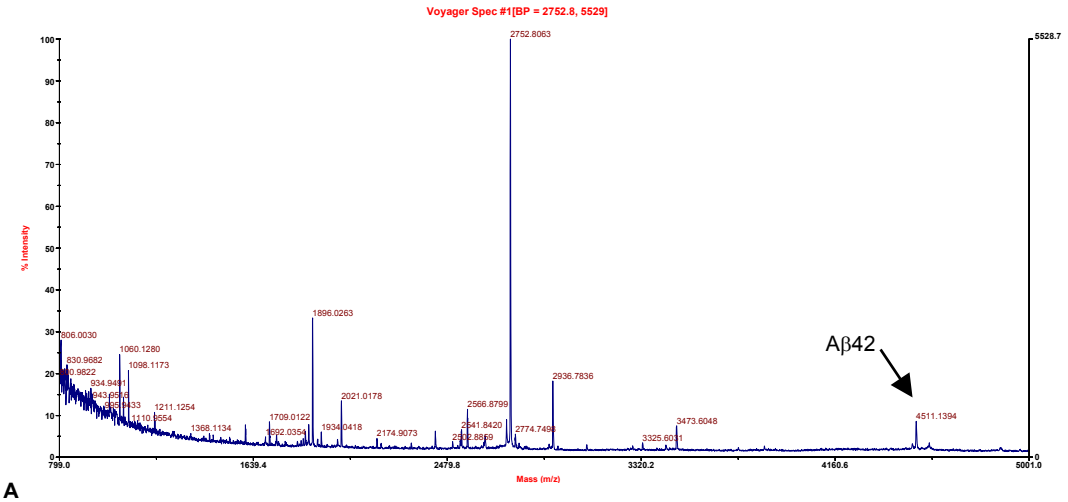
8.2.2.2 MALDI-TOF MS of the medium mass range

The medium size mass range consisted of proteins of m/z between 5 kDa and 100 kDa. This is believed to be the optimum mass range for the MALDI-TOF technique.

Although no internal calibrators had previously been quantitated in plasma for this mass range, the product of the alternatively spliced form of *CFH*, factor H-like 1 (FHL1) protein resided in this window. With a mass of approximately 50 kDa, quantitative assessment of this protein in plasma would be of great interest, especially in relation to the full length CFH protein, with a size of 150 kDa.

Ten microlitres of crude plasma was used as starting material. This was diluted and subject to immunodepletion as described (Section 2.4.8.1) in three independent but parallel experimental replicates. Following the depletion of high abundance proteins, the flow-through containing the low abundance proteins was subjected to hydrophobic interaction chromatography using C18 magnetic beads, as described in Section 2.4.8.2.

The results showed a significant improvement in the number of peaks detected and in their intensities (**Figure 8.6**). Comparison of the resulting spectra with the mass spectrum of human crude plasma showed a significant reduction in the intensity of the serum albumin peak. Furthermore, the level of background noise particularly in the lower m/z range was greatly reduced after treatment with C18 magnetic beads.



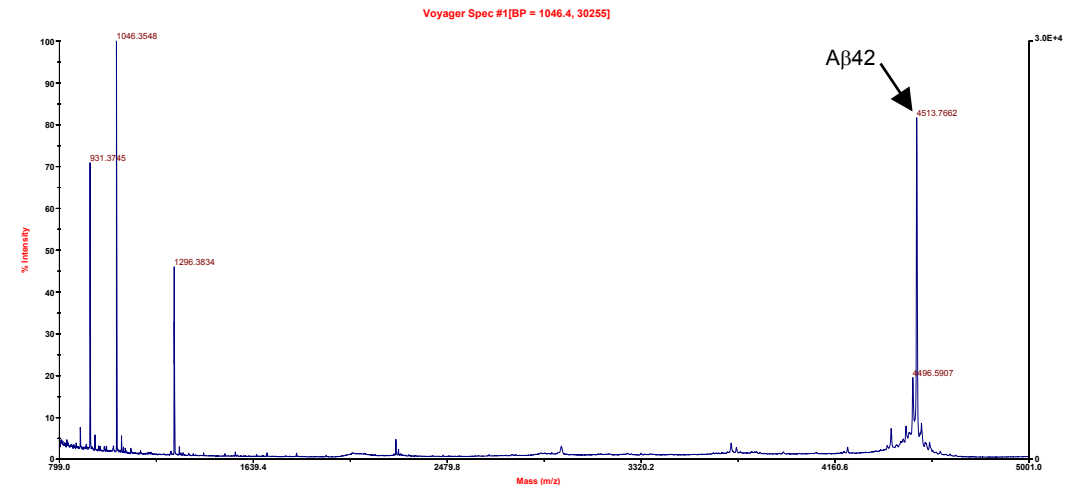


Figure 8.5 Assessment of the MALDI-TOF technique in quantitative determination of differences in Aβ42 concentrations after ZipTip™ treatment.

MALDI-TOF MS analysis of plasma samples spiked with detectable concentrations of the human Aβ42 peptide showed large variation between replicates (data not shown) and no correlation between different Aβ42 concentrations and peak intensities. Synthetic Aβ42 peptides of 500 ng /ml **(A)** and 1000 ng/ml **(B)** concentration were desalted and concentrated by treatment with ZipTip™. Resulting spectra were analysed by co-elution with alpha-cyano matrix and spotting onto a MALDI plate followed by detection on MALDI-TOF MS Voyager-DE™ STR (Applied Biosystems) in a linear mode.

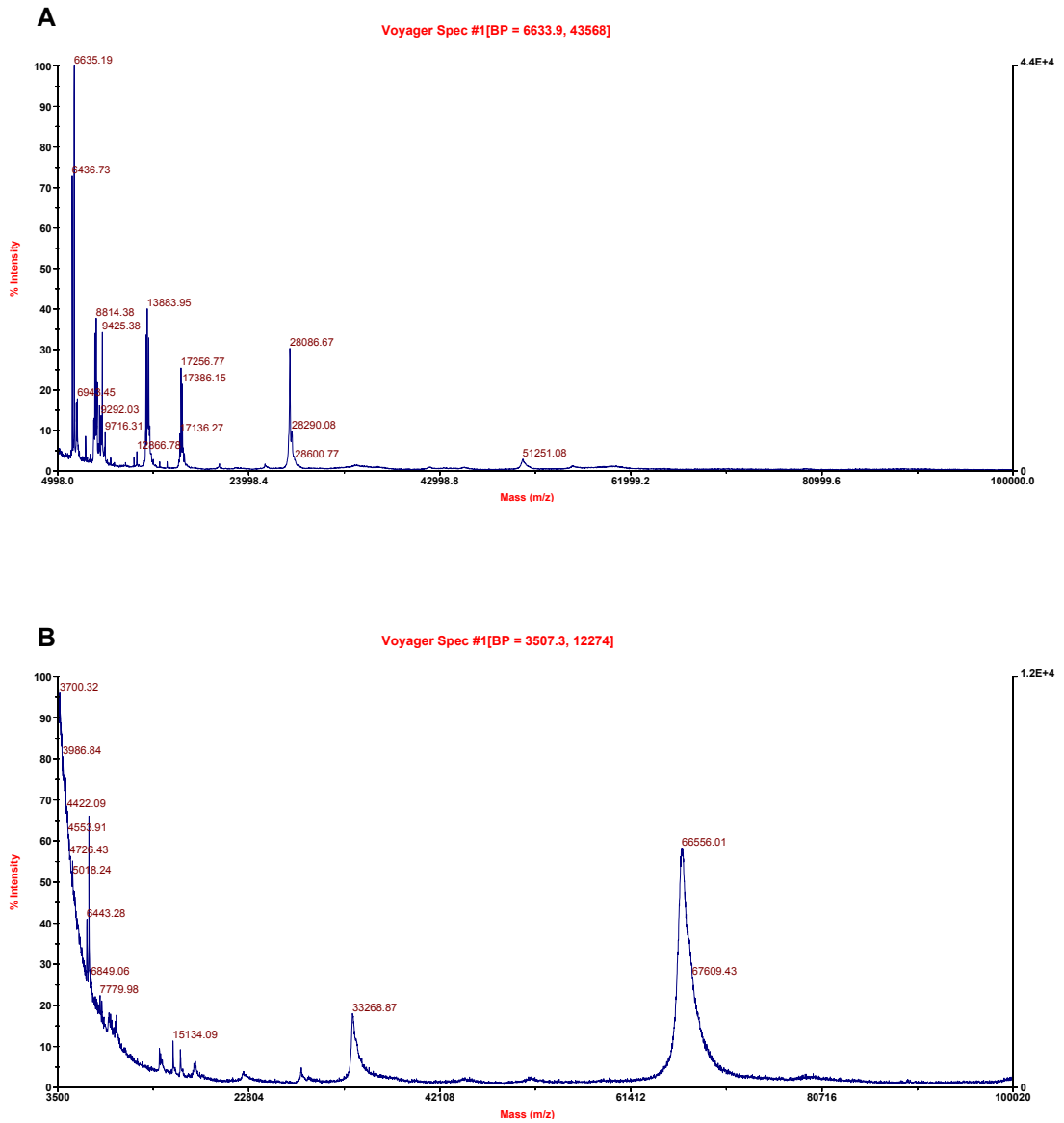


Figure 8.6 MALDI-TOF MS results of the human plasma medium size mass range after immunodepletion and hydrophobic interaction chromatography.

Mass spectrum of the human plasma in the medium size range after immunodepletion and hydrophobic interaction chromatography showing an increase in the signal:noise ratio and the number of detectable peaks after treatment. Mass spectra were obtained for the medium size mass range (4.9-100 kDa). **A.** Treatment of human plasma by immunodepletion and magnetic bead separation using C18 magnetic beads resulted in reduced background noise and larger number of detectable peaks. **B.** The mass spectrum of crude human plasma, of the medium mass range, is shown for comparison. The significance value, on the right axis, shows the level of confidence with which signal intensities were detected which also showed an improvement compared with the pre-treatment techniques used before. Spectra were obtained using the MALDI-TOF MS Voyager-DE™ STR (Applied Biosystems) in a linear mode.

A number of peaks were also observed at m/z regions of 6, 8, 13, 17 and 28 kDa. Another feature of the improvement in signal detection was the significance at which the peak intensities were observed by the instrument. The significance value reached 4.4E+04 for the highest peak observed at m/z of 6,635. The significance of a mass peak is a function of the number of ionised species detected from a precursor protein at each m/z signal.

8.2.2.3 MALDI-TOF MS of the large mass range

The large mass range proved to be the most challenging due to the difficulty in exciting molecules of large mass to ionisation (Geno & MacFarlane, 1989). On the other hand, proteins belonging to the small mass range, despite ionisation being easier, tended to disappear in the time of flight (TOF) chamber and were not detected.

Several methods were assessed in the analysis of the large mass range, including immunodepletion, magnetic bead separation as well as C4 ZipTip™ chromatography specially designed to aid detection of proteins of large mass. MALDI-TOF MS results after dilution, acidifying and ZipTip™ treatment of plasma samples did not significantly differ from those using immunodepleted plasma (**Figure 8.7**). Analysis of the large mass range using crude plasma was not possible due to the high level of interference by large, high abundance proteins which disrupted the signal. This occurred due to ionisation attempts with the large molecular weight fraction of plasma. The spectrum of immunodepleted plasma showed a number of peaks at m/z of 131,970, 143,090 and 161,764 with a high level of background noise. Two large regions of m/z of approximately 132,804 and 147,509 were also apparent in the immunodepleted spectrum but were almost lost in the background noise. The intensity of the peak at m/z of approximately 134,079 increased dramatically from 50% to 100% (**Figure 8.7**, left y-axis) despite low confidence (358, **Figure 8.7-A**; 812, **Figure 8.7-B**), which accounted for the low number of ionised protein species being detected from the sample.

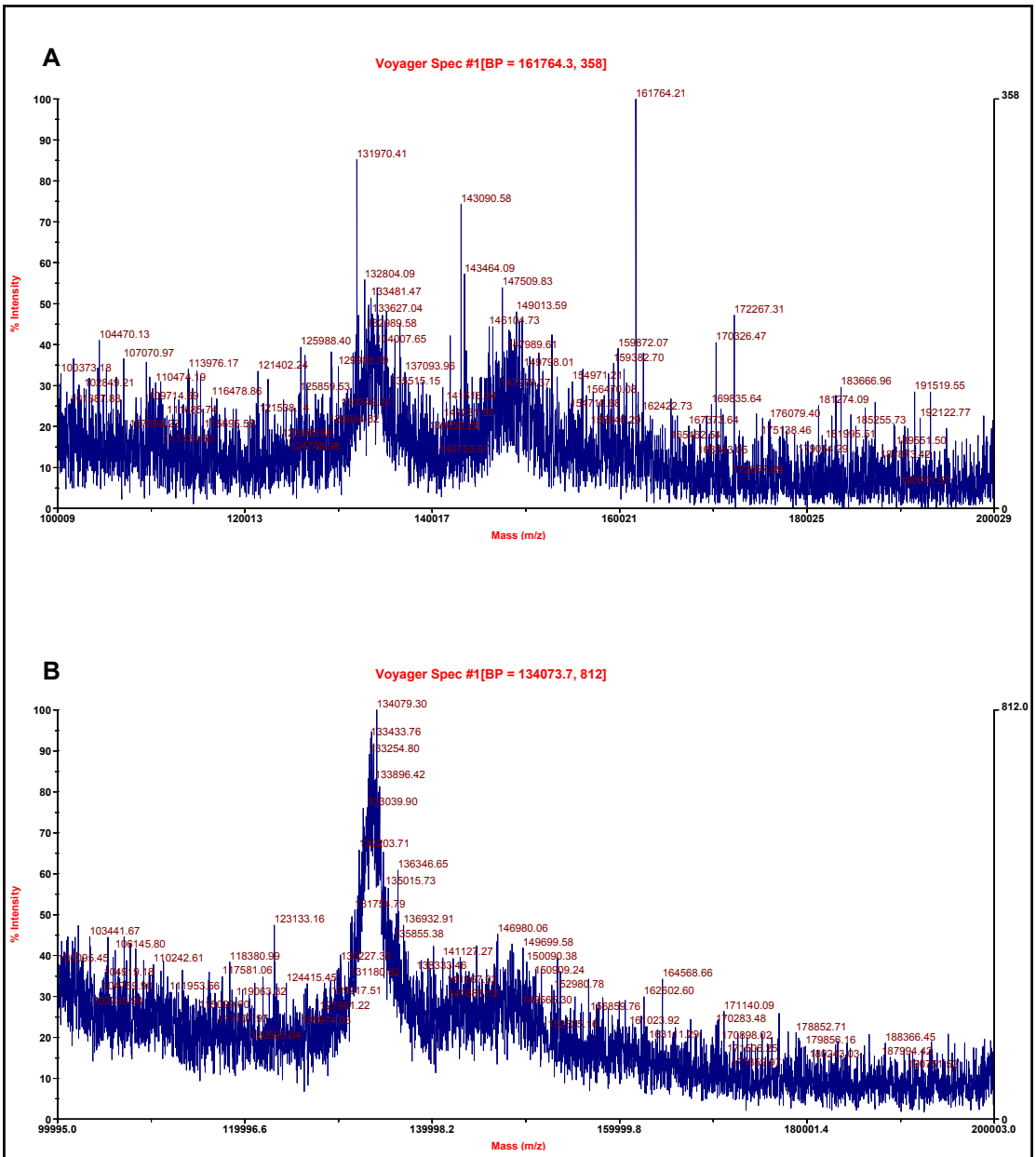


Figure 8.7 MALDI-TOF MS results of the large mass range after immunodepletion and ZipTip™ treatment of crude human plasma.

Mass spectra of the human plasma after immunodepletion and ZipTip treatment showing the difficulty associated with analysing the large mass range (100-200 kDa) due to the disrupted signal and high background noise believed to be caused by interference from large and high abundance proteins. **A.** Immunodepleted human plasma showing a high degree of complexity and lack of sufficient ionisation, as estimated by the low significance value on the right axis (see text). **B.** The spectrum of diluted, acidified and concentrated human plasma, using C4 ZipTip™ treatment specifically targeted at the larger fraction of plasma, not showing sufficient improvement in signal strength or confidence. Spectra were obtained using the MALDI-TOF MS Voyager-DE™ STR (Applied Biosystems) in a linear mode.

8.2.3 Capillary electrophoresis

Capillary electrophoresis (CE) techniques were used primarily for the quantitative separation and detection of protein samples by UV absorbance. DL-tryptophan was initially used as a surrogate for a peptide in order to optimise CE techniques for sample concentration and injection time. The tryptophan molecule absorbs UV light at 280 nm and can therefore be detected by CE. The aim was to use optimised CE conditions for quantitative measurements of protein concentrations in complex mixtures including plasma. Thereafter, the CE fractionation phase could be coupled to a 12T FT-ICR instrument for ionisation followed by fragmentation and detection by MS.

8.2.3.1 Optimisation of peptide concentration

Three concentrations (10 μM, 100 μM and 1,000 μM) of the amino acid tryptophan, diluted in the CE Running Buffer, were used in the initial analysis. The injection duration time was set at 12 seconds. Negative controls, consisting of the running buffer alone, were included in each run.

Results of the CE fractionation of plasma were found to be highly consistent (**Figure 8.8**) and peak areas were significantly correlated with peptide concentration. Each sample was treated and analysed using three replicates in parallel. Mean peak areas for each sample, observed at 100 μM and 1,000 μM concentrations of tryptophan, were found to be 23.6 absorbance over time (A/t) (± 0.2 SD) and 250.6 A/t (± 2.7 SD), respectively (**Figure 8.8**). The mean peak area observed for the 10 μM concentration of tryptophan was too low and classified as background noise. However, manual inspection of the graph showed the presence of a small peak at 5.8 minutes (**Figure 8.8**). Low levels of inter-sample variation were observed (CV < 0.05). Amino acid peaks were also consistently detected at 5.9 minutes (± 0.2 SD) in each experiment which was used as a measure of the reproducibility of CE.

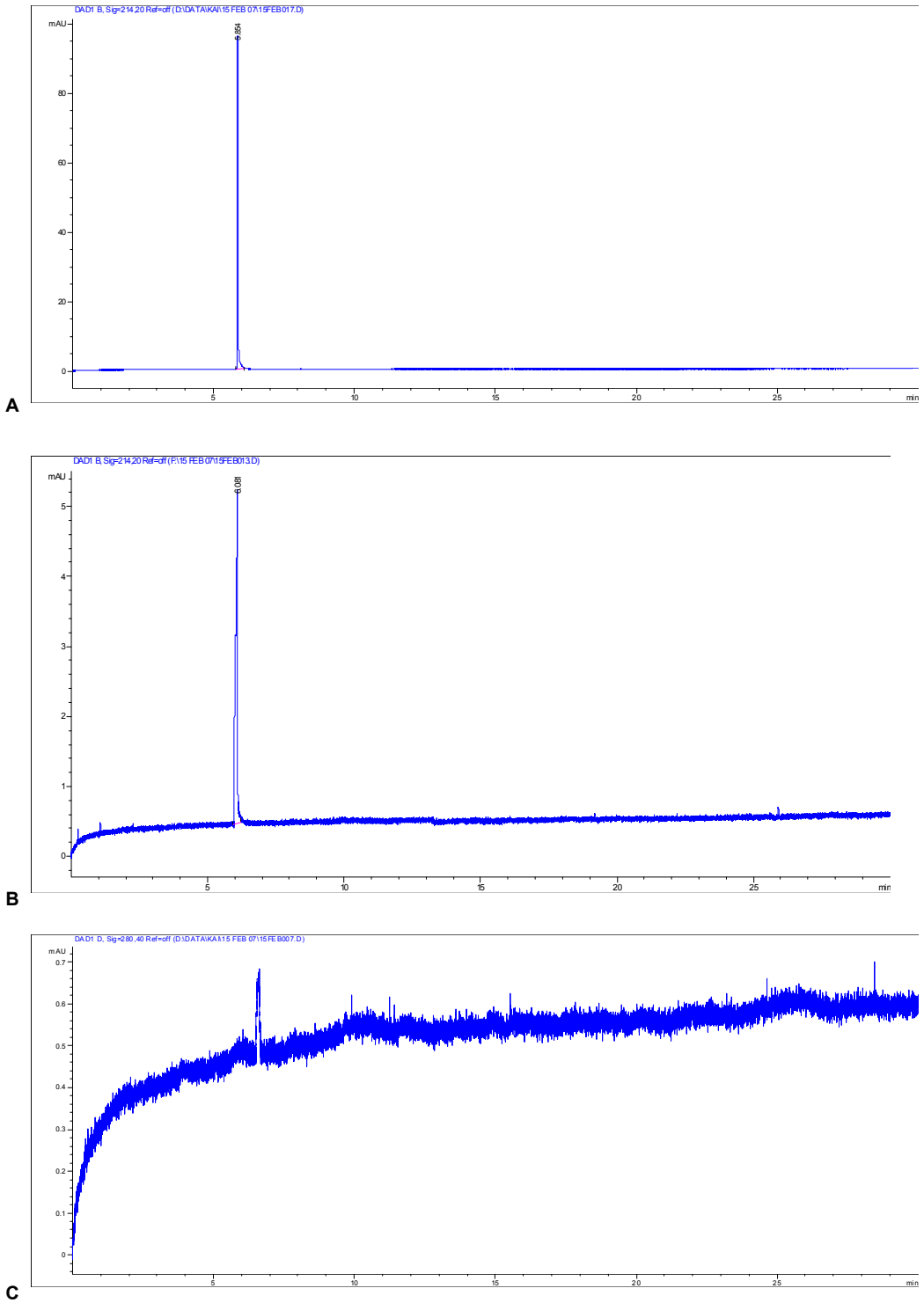


Figure 8.8 Results of capillary electrophoresis of varying concentrations of tryptophan.

UV detection of tryptophan at three different concentrations using CE showed that peak areas were highly correlated with sample concentration. UV absorbance was detected at 280 nm for 10 μM **(A)**, 100 μM **(B)** and 1000 μM **(C)** concentrations of DL-tryptophan, diluted in HPLC water. The peak of DL-tryptophan was consistently detected at 5.9 minutes (±0.2, SD). An injection time of 12 seconds was used and each graph is the result of three replicates (CV < 0.05). Negative controls, consisting of the running buffer only, were used in each experiment. Live acquisition was carried out using the Agilent ChemStation CE instrument.

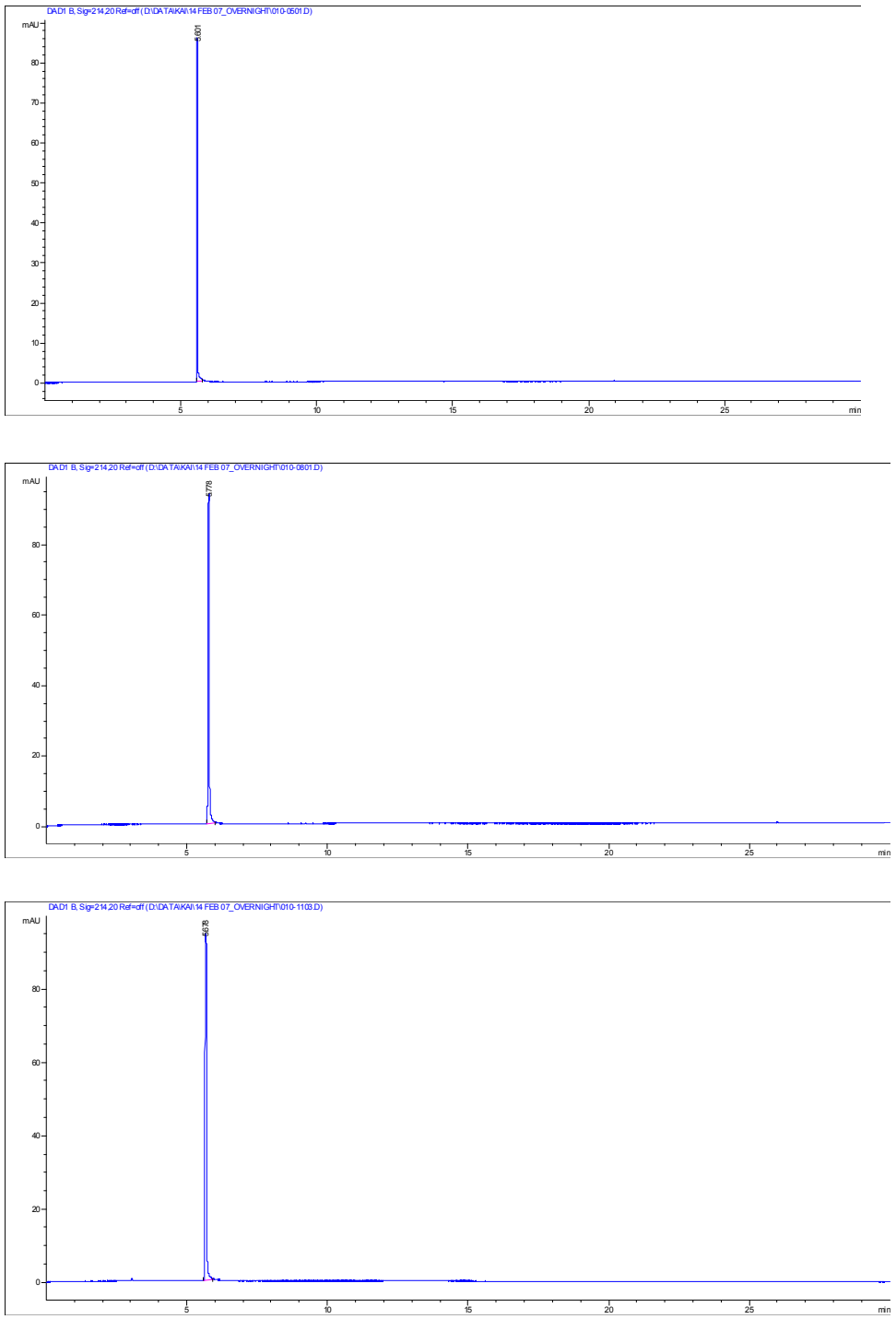


Figure 8.9 Capillary electrophoresis results of varying the length of injection time on the observed tryptophan peak area.

UV detection of 1mM DL-tryptophan at three different sample injection times showed a direct and consistent relationship between the injection time and tryptophan peak area. DL-tryptophan was diluted in HPLC water and used to assess the effect of different injection times of 6 seconds **(A)**, 12 seconds **(B)** and 18 seconds **(C)** on the tryptophan peak areas. Each analysis was carried out in triplicate and included a negative control, containing the running buffer only. The tryptophan peak was consistently detected at 5.7 (±0.1, SD) minutes. Live acquisition was carried out using the Agilent ChemStation CE instrument. For further discussion see text.

8.2.3.2 Optimisation of injection times

The 1 mM concentration of DL-tryptophan resulted in the least amount of variation between replicates and the highest resolution (**Figure 8.8**). This concentration of tryptophan was therefore used to optimise the duration of injection time by comparing the results at 6, 12 and 18 seconds.

The results showed a direct and consistent relationship between the injection time and the peak area observed for 1 mM of tryptophan (**Figure 8.9**). Mean peak areas varied from 128.0 absorbance over time (A/t) (± 1.1 SD) at 6 seconds length of injection to 254.9 A/t (± 1.7 SD) at 12 seconds and 368.8 A/t (± 2.4 SD) at 18 seconds. No absorption peaks were observed in the negative control sample which consisted of the running buffer alone. At all three injection times, peaks of tryptophan were observed consistently at the mean time of 5.7 A/t (± 0.1 SD) minutes indicating good reproducibility. Low levels of inter-sample variation were observed between replicates (CV < 0.05).

Tryptophan was used as a surrogate peptide for UV detection by CE which showed a potential for quantitation of simple molecules based on their UV absorbance. The results of the varying sample concentration and injection times were promising as they showed a direct and consistent relationship between the two parameters of sample concentration and injection time, and the peak area observed. The optimised CE conditions could therefore be applied to single protein samples in order to assess the CE techniques in handling protein samples.

8.2.3.3 CE analysis of single protein samples

Following the optimisation of amino acid (tryptophan) concentration and injection times, analytes containing a single protein were analysed using the protocol described above. Initially, three proteins consisting of ubiquitin, myoglobin and BSA were used individually for CE, at two different concentrations (10 μg/ml and 50 μg/ml).

The results of the assessment of CE for quantitation of single protein samples of known concentration showed a high level of variation between replicates (CV > 0.05). In the case of 10 μg/ml ubiquitin, no peaks were observed due to the high

level of background noise in all three replicates. Myoglobin (10 μg/ml) showed a high level of inter-experimental variation (CV = 26.1%). BSA (10 μg/ml) also showed a high level of variation between replicates (CV = 24.6%). In the case of 50 μg/ml ubiquitin, even higher levels of inter-experimental variation were observed between measurements (CV = 76.1%). A similar pattern was seen for myoglobin (50 μg/ml) and BSA (50 μg/ml) samples with observed CVs of 78.6% and 82.0%, respectively. Protein peaks were detected at the same time points for all three proteins, despite their molecular weight differences. The same peaks were also detected in the negative control samples containing the running buffer alone suggesting contamination. At the 10 μg/ml concentration, peaks of myoglobin and BSA were detected at 50.8 minutes (1.49, SD) and 53.4 minutes (0.42, SD), respectively. At the 50 μg/ml concentration, peaks of ubiquitin, myoglobin and BSA were observed at 41.1 minutes (4.34, SD), 41.5 minutes (0.56, SD) and 42.6 minutes (0.68, SD), respectively. Troubleshooting procedures, described in Methods Section 2.4.8.6.2, were employed to eliminate the problem of high background noise (possibly the result of air bubbles in the solutions) and contaminating peaks detected in the CE spectra, which may have been the result of contaminated buffers or un-dissolved proteins.

The findings of the single protein analysis by CE showed high levels of variation between replicates and a lack of consistency between observed peaks. The trouble shooting procedures aimed at eliminating the contaminating peaks observed in the negative control samples resulted in loss of UV detection. Although the preliminary experiments assessing the quantitative abilities of CE involving tryptophan were promising, UV detection of samples of single proteins was unsuccessful. The detection and analysis of proteins by CE coupled with FT-ICR MS was then assessed, despite the problem with detection of protein samples by UV, as discussed below.

8.2.3.4 CE coupled with FT-ICR MS

CE was coupled with FT-ICR mass spectrometry for the detection and quantitation of single protein samples containing ubiquitin (8.5 kDa), myoglobin (16.7 kDa) and

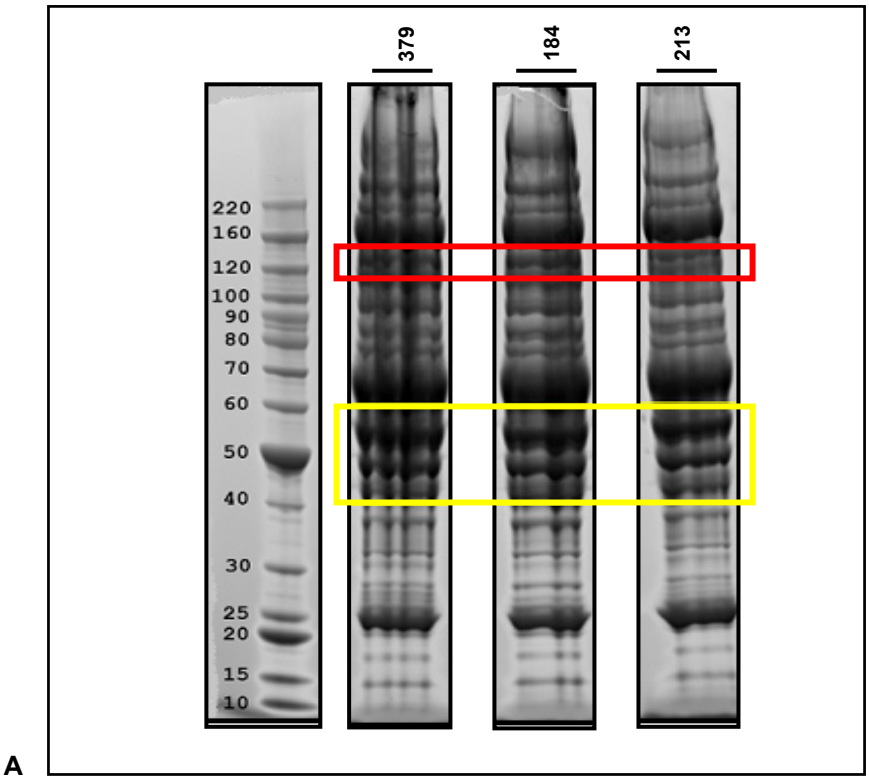
BSA (66.4 kDa) using the protocol described in Section 2.4.8.6. Negative controls containing the CE Running Buffer alone were used in each experiment. MS detection was performed on a 12T quadrupole system APEX-QE (Bruker Daltonics, Germany) equipped with a Fourier transform mass spectrometer.

Sample fractionation by CE followed by detection with FT-ICR mass spectrometry was carried out several times with single protein samples containing ubiquitin, myoglobin and BSA. No protein peaks were detected from the analysis, and mass spectra mainly consisted of peaks of buffer components of small m/z. The detection ability of the quadrupole instrument was tested by directly injecting 1 mg/ml of myoglobin diluted in HPLC H_2O into the electrospray chamber. Peaks of protonated myoglobin ions were detected, representing the 16.9 kDa ionised protein. However, attempts to detect protein samples of known concentration, using CE FT-ICR MS were unsuccessful.

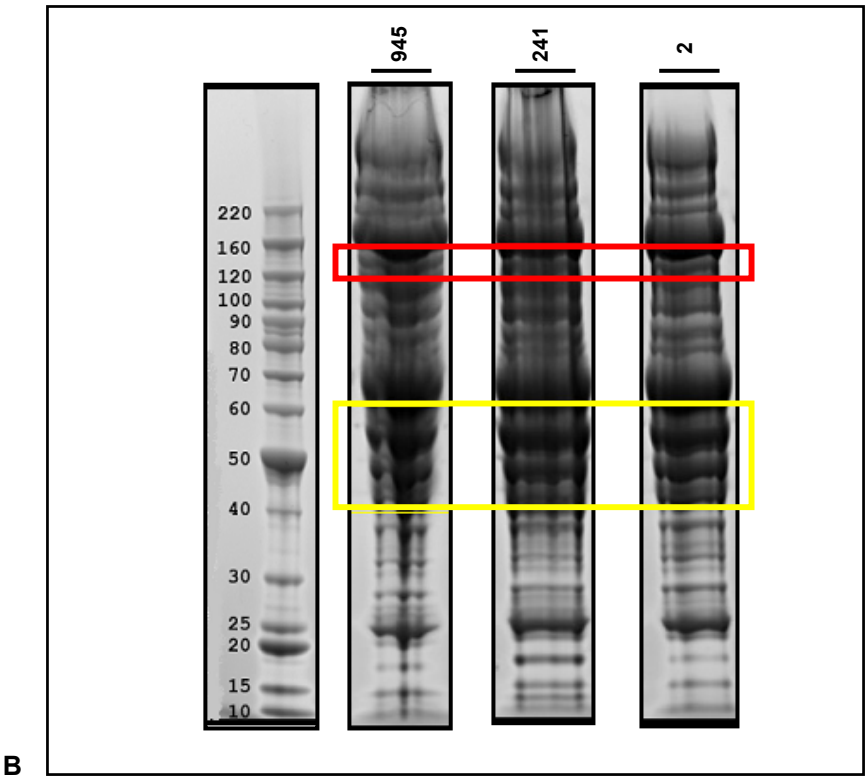
8.2.4 Ion exchange chromatography

Three Croatian plasma samples were selected from the higher and lower ends of the CFH distribution as measured by ELISA. Samples were immunodepleted then separated by 4-12% pre-cast NuPAGE® gel electrophoresis. Gel bands corresponding to the 150 kDa full length CFH (**Figure 8.10**, red box) and the 55 kDa factor H-like protein (**Figure 8.10**, yellow box) were excised, trypsin-digested and analysed by ion exchange chromatography (IEC) followed by nLC-MS/MS. IEC and nano LC MS/MS experiments were conducted by Mr. Douglas Lamont at the University of Dundee. The aim was to assess the correlation between plasma CFH concentrations measured by ELISA and IEC coupled with nLC-MS/MS. This also provided the opportunity to analyse plasma levels of the 55 kDa factor H like protein in relation to the 150 kDa full length CFH.

Analysis of the lower molecular weight fraction was unsuccessful due to lack of detection of peptides corresponding to the factor H-like 1 (FHL1) protein. Peptide ions were identified from the larger molecular weight fraction using the protein search engine Mascot. Five peptide species with the highest ion scores were then extracted from each sample.



A



B

Figure 8.10 Plasma protein staining prior to ion exchange chromatography.

Results of the protein staining of plasma samples selected for ion exchange chromatography. Three immunodepleted plasma samples of 230-250 μg/ml **(A)** and 590-620 μg/ml **(B)** CFH concentration were separated on 4-12% pre-cast NuPAGE® Novex Bis-Tris gels and stained using SimplyBlue™ SafeStain. Gel regions containing the 150 kDa full length CFH protein (red box) and the 55 kDa factor H-like1 protein (yellow box) were excised and analysed. Sample identification numbers are shown above each well.

The corresponding ion peaks were identified on the mass spectra based on the retention time of the ions. Peak areas corresponding to CFH peptide ions were normalised to plasma CFH levels, and compared between samples of high CFH (samples IDs 945, 241 and 2) and low CFH (samples IDs 379, 184 and 213) concentration (**Figure 8.11**). In four out of the five peptide ions detected, peak areas were found to vary considerably between the six samples and did not correlate significantly with plasma CFH concentrations ($R^2 < 0.05$, $P > 0.10$) using a linear regression analysis of variance (**Figure 8.12 A-D**). One of the five selected peptide ions was found to significantly correlate with findings of the CFH ELISA ($R^2 = 0.954$, $P < 0.01$) (**Figure 8.12-E**). The overall squared correlation was also calculated for the combined set of data but no significant correlation was observed between IEC and ELISA measurements of plasma CFH ($R^2 = 0.05$, $P > 0.10$).

The results of the IEC experiments coupled with nLC-MS/MS showed the problems of reliably and reproducibly determining plasma CFH concentrations when compared with measurements of ELISA which was successful in only one out of five experiments.

8.3 Summary and discussion

The results of assessing proteomic techniques such as MALDI-TOF MS, CE FT-ICR MS and IEC nLC-MS/MS, in protein identification and quantitation were presented in this chapter. The aims were to correlate measurements of plasma proteins performed by ELISA with the MS results, and to assess the potential of MS-related techniques in biomarker discovery and quantitation.

Initially, immunodepletion of high abundance plasma proteins, albumin, IgG, IgA, transferrin, haptoglobin and α_1-antitrypsin, using a multiple affinity column, was tested using MALDI-TOF MS, which proved successful in improving the MS signal (**Figure 8.1**). Immunodepletion by multiple affinity removal columns is an effective method of removal of highly abundant proteins simultaneously, with minimal carryover and low non-specific binding (Bjorhall *et al*., 2005).

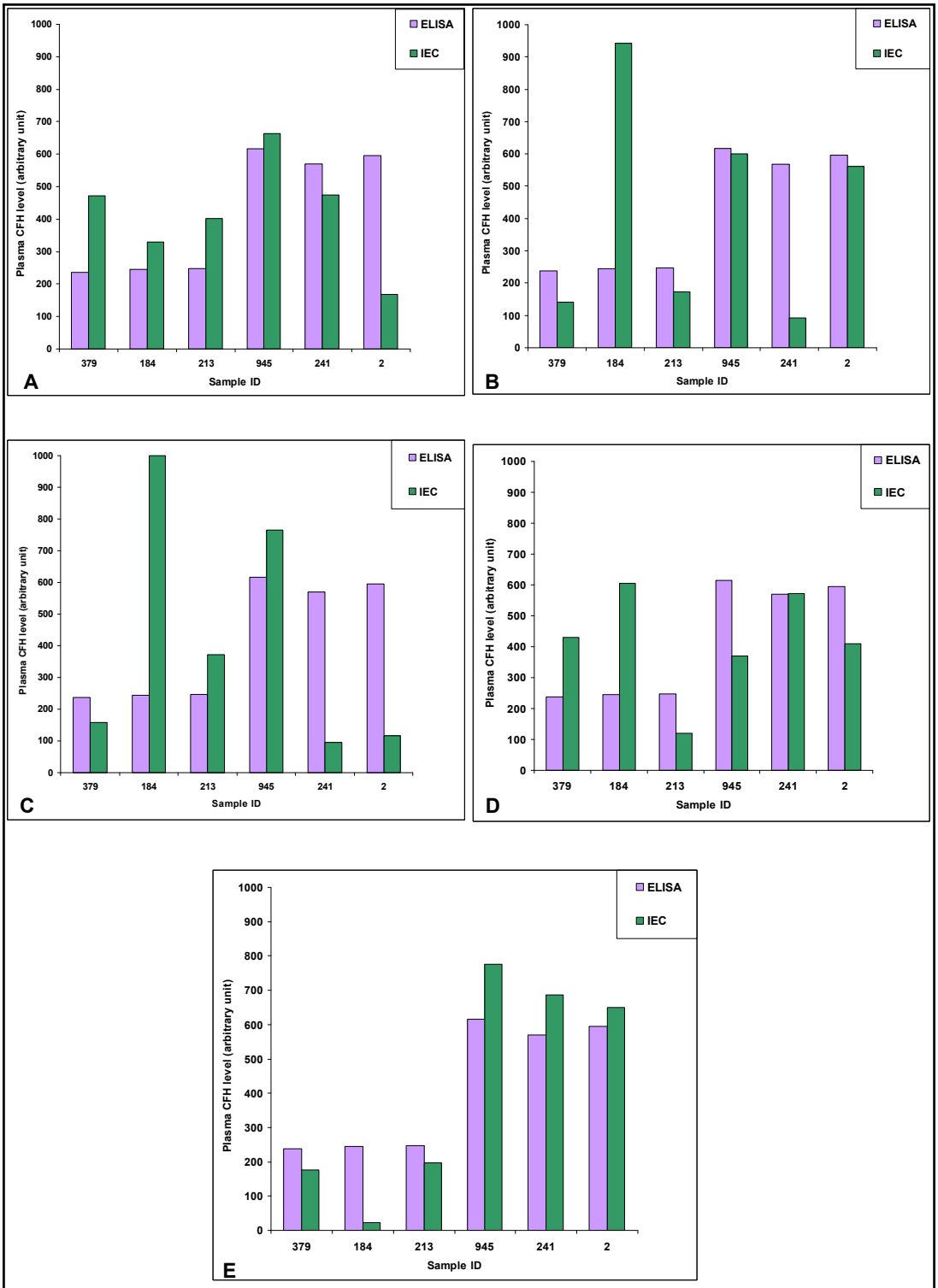


Figure 8.11 Comparison of CFH concentrations measured by ELISA and IEC.

Significantly correlated measurements of plasma CFH were observed in some peptide fractions as compared by ion exchange chromatography. Peak areas from six immunodepleted plasma samples were measured for CFH peptide ions produced by ion exchange chromatography (IEC) and detected by nLC-MS/MS. Estimated measurements were compared to CFH concentrations determined by ELISA. Five independent peptide ions were analysed as shown in A-E. Peak areas were normalised to CFH concentration (μg/ml) in each sample.

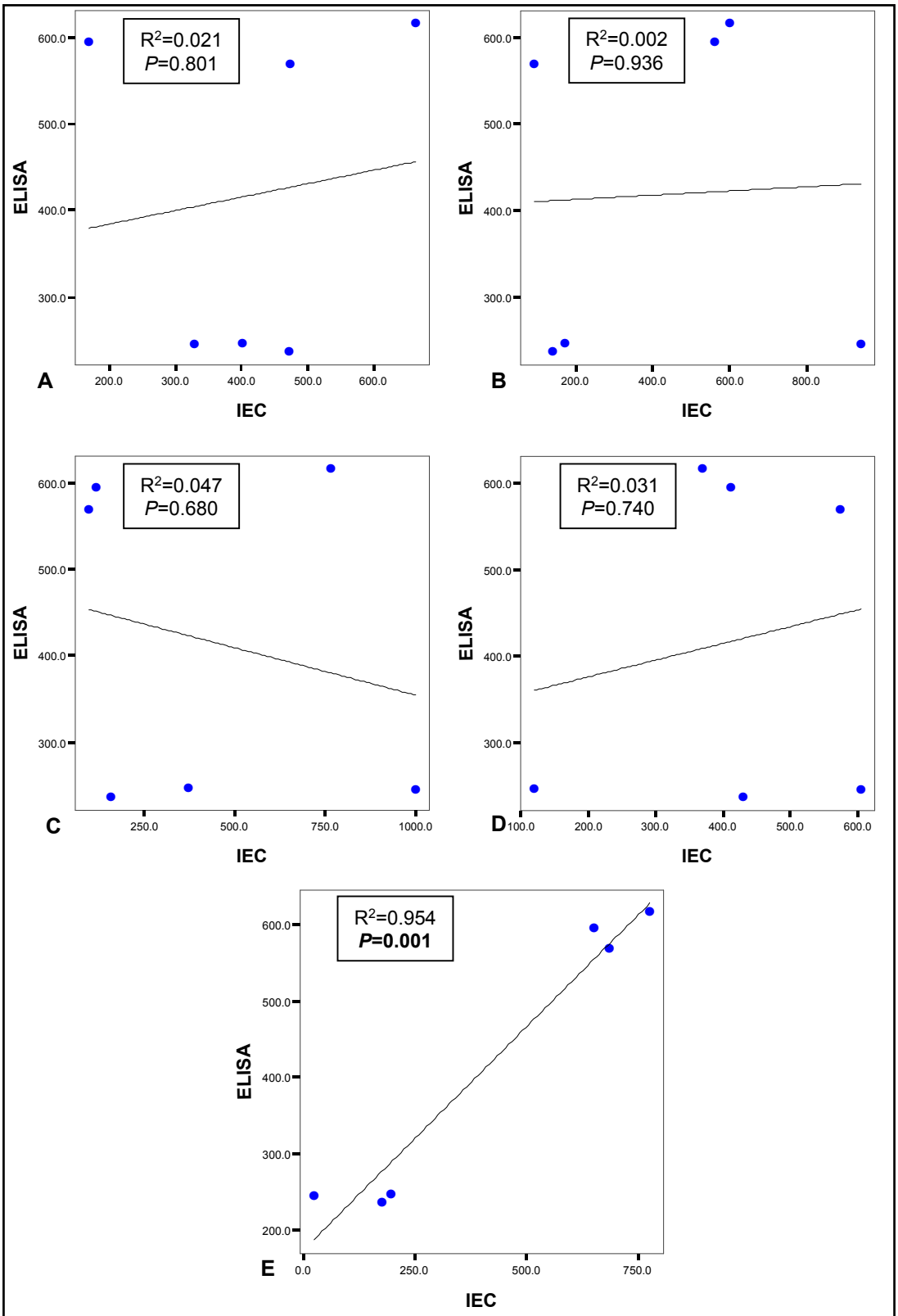


Figure 8.12 Test of correlation between ELISA and IEC measurements of plasma CFH.

Squared correlation coefficients (R^2) and their significance (*P*) were obtained using linear regression analyses, comparing the measurements of plasma CFH obtained by ELISA and IEC. Results showed that one out of five peptide ions correlated significantly with plasma CFH measurements of ELISA in all six plasma samples (**E**). Analysis was carried out using IEC normalised peak areas obtained for the five independent peptide ions (**A-E**) shown in Figure 8.11, using the software SPSS v.12.

Although depletion of highly abundant proteins in plasma significantly increases the number of proteins identified, studies have shown that both the degree of sample complexity and the depletion method result in a non-selective loss of other proteins of lower abundance, such as cytokines (Yocum *et al*., 2005; Granger *et al*., 2005). An example relevant to the present study is the claim that greater than 95% of Aβ40 and Aβ42 in human plasma reside within a sequestered pool of high abundance proteins such as albumin and transthyretin (Kuo *et al*., 1999; 2000). A possible approach to address this problem is to disrupt the binding of the low molecular weight proteins to large carrier proteins by using denaturing conditions such as addition of acetonitrile to the plasma sample (Huang *et al*., 2005).

MALDI-TOF MS analysis of plasma was performed in three protein mass ranges of small (1 kDa < m/z < 5 kDa), medium (5 kDa < m/z < 100 kDa) and large (100 kDa < m/z < 200 kDa) size. Firstly, analysis of the small mass range, after treatment of crude plasma with ZipTip resulted in a significantly improved signal compared with immunodepletion, and a number of reproducible MS peaks were detected at m/z of 1,896, 2,937 and 3,437 (**Figure 8.2**). Ultrafiltration of plasma, using 10 kDa molecular cut-off filters, followed by treatment by ZipTip resulted in even higher signal:noise ratio, probably due to the larger volume of starting plasma sample (**Figure 8.3**). However, spiking the samples of human plasma with known concentrations of purified synthetic Aβ42 showed the lack of sensitivity associated with MALDI-TOF MS in detecting samples containing less than 500 ng/ml, which is considerably higher than the physiological concentrations of Aβ42 in the human plasma (**Figure 8.4**). Furthermore, the 500 ng/ml and 1,000 ng/ml concentrations of spiked Aβ42 at m/z of 4.5 were poorly correlated with each other, and high levels of inter-sample variation (CV > 20%) were observed between replicates (**Figure 8.5**).

Secondly, analysis of the medium mass range of plasma proteins, after immunodepletion and separation by hydrophobic interaction chromatography using magnetic beads, resulted in higher signal:noise ratio and improved signal intensities (intensity of 4.4E0+04 for the highest peak at m/z of 6635) (**Figure 8.6**). The aim was to detect the 55 kDa factor H-like protein in plasma, however this was unsuccessful and attempts at improving the spectra for mass range of 40 kDa and 100 kDa, by altering ionisation intensities and trying other techniques such as desalting

and concentrating samples, did not result in improved sensitivity. Magnetic bead separation techniques have been shown to greatly improve plasma peptide detection by MALDI-TOF MS (Villanueva *et al*., 2004). Studies have reported peptide marker profiles detected using magnetic bead separation followed by MALDI-TOF MS in plasma samples of individuals affected by cancer compared with controls (Villanueva *et al*., 2006; van der Werff *et al*., 2008). However, such studies have been restricted to pattern recognition between disease and control samples and have not involved quantitative assessments of protein profiles.

Thirdly, analysis of the large mass range after immunodepletion resulted in a number of detectable peaks at m/z of 131, 143 and 167 kDa, but only in the presence of high levels of background noise. Immunodepletion followed by C4 ZipTip treatment resulted in slightly improved spectra, with two regions of approximately 134 and 147 kDa becoming visible, despite the low intensity of peaks (**Figure 8.7-B**).

Despite its widely accepted use in identification and characterisation of large molecules (Karas & Hillenkamp, 1988; Henzel *et al*., 1993, 2003), the results of the MALDI-TOF mass spectrometry in the present study were highly variable between replicates. The high level of variation can be due to experimental limitations such as heterogeneous distribution of sample molecules on the MALDI spot. This results in large variation between samples at different locations on the spot, which would create a different number of ionised molecules. This is further complicated since the crystallisation with the matrix does not occur uniformly on each spot. Considering the above caveats and the number of "upstream" experimental procedures carried out prior to mass spectrometry, it is not unexpected that high levels of inter-experimental variation was observed in MALDI-TOF MS. In order to reduce the inter-experimental variation, the instrument should be calibrated using test samples of known size and concentration or use of internal standards.

Capillary electrophoresis (CE) was used as a method of fractionation of proteins in the sample prior to MS detection and quantitation. Initially, UV detection and quantitation of single peptide samples was assessed, using an optimised protocol (Weissenger *et al*., 2004). Firstly, optimisation of peptide concentration, using 100 μM and 1000 μM of the amino acid tryptophan, resulted in peak areas of 23.6 A/t (±

0.2 SD) and 250.6 A/t (± 2.7 SD), respectively (**Figure 8.8**). Peak areas were found to be highly correlated with sample concentrations and between replicates (CV < 5%). These results were promising and they showed that the specificity of CE was enough to detect 10-fold differences in concentration of a single amino acid.

Secondly, optimisation of injection duration times, using 1 mM tryptophan, resulted in mean peak areas which showed a strong correlation with length of injection time for the single amino acid tryptophan (**Figure 8.9**). Furthermore, peaks of tryptophan were observed consistently at the mean time of 5.7 minutes (± 0.1 SD), indicating good reproducibility of results.

Tryptophan was used in the CE optimisation procedures since it is a simple amino acid which readily absorbs UV light. Optimised concentration and injection times were then applied to three single protein samples, ubiquitin, myoglobin and BSA. However, the results showed high levels of variation between replicated measurements (CV > 5%). Furthermore, observed peaks were believed to be the result of contamination, since all detectable peaks occurred at the same time-point, despite differences in molecular weights of the proteins. However, the problem was not solved by the troubleshooting procedures described before.

Some studies have succeeded in differentially identifying more than a thousand polypeptide species in samples of diseased and control individuals using CE coupled with mass spectrometry in plasma and urine (Kaiser *et al*., 2004; Sassi *et al*., 2005). Notably, a study involving the comparison of polypeptide patterns between urine samples obtained from healthy and sick individuals affected by a number of kidney diseases identified isoforms of albumin which significantly differed in "levels" between cases and controls using CE coupled with FT-ICR MS (Weissinger *et al*., 2004). In the present study, mass spectrometric detection of protein samples after fractionation with CE was assessed after coupling the CE instrument to an FT-ICR mass spectrometer. Analysis of three single protein samples consisting of ubiquitin, myoglobin and BSA using CE coupled with FT-ICR followed by mass spectrometry resulted in detection of no sample peaks. Repeated attempts at detecting protein samples by FT-ICR MS resulted only in detection of low molecular weight buffer components and so were unsuccessful.

Following the assessment of the two "top-down" mass spectrometry approaches which involved the analysis of intact proteins, a "bottom-up" approach was used involving fractionation by ion exchange chromatography (IEC) followed by trypsin digestion and nLC-MS/MS analysis. Experimental procedures were carried out using plasma samples selected from the higher and lower ends of the distribution of CFH concentration. Six samples were immunodepleted, separated on 4-12% pre-cast polyacrylamide gels and stained (**Figure 8.10**). Analysis of two molecular weight fractions, excised from the stained gels, containing the 150 kDa full length CFH and the 55 kDa factor H-like proteins, resulted in detection of full-length CFH but failed to detect the shorter factor H-like 1 protein due to its lower abundance (believed to be about 10 times lower than full length CFH, ranging between 10-50 μg/ml). Five peptide ions of high intensity were selected from the higher molecular weight fraction (**Figure 8.11 A-E**) in which the intensity of the peak representing the full length protein was measured. In one of the five selected peptide ions, the concentration of the ion exchange chromatography peaks were found to correlate significantly with those of the CFH ELISA ($R^2 = 0.954$, $P < 0.01$) (**Figure 8.12-E**). The overall squared correlation of all the five peptides obtained from the six plasma samples and compared between IEC and ELISA, did not differ significantly from zero ($R^2 = 0.05$, $P > 0.10$). However, these results were more promising as far as the quantitation of plasma proteins is concerned, compared with the "top-down" approaches discussed before. The results of the IEC experiments coupled with nLC-MS/MS showed that in one out of five experiments, plasma CFH measurements were highly correlated with measurements obtained by ELISA.

The results of the mass spectrometry experiments highlighted several shortfalls of the proteomics techniques in relation to protein identification and quantitation. Firstly, the most common issues are abundance and interference. The results of the present study showed that interference, not only from the most common plasma proteins such as albumin and immunoglobulins, but from any protein species, which are often orders of magnitude more abundant than the protein of interest, can limit the detection by MS. Data obtained using IEC nLC-MS/MS, which is a highly sensitive method of protein detection, confirmed the above result by the failure to detect the short Factor H-like 1 protein in a mixture containing proteins of

approximately 150 kDa and 50 kDa after separation by electrophoresis. Indeed, most "biomarkers of disease" which have been identified through protein profile differences seem to be of relatively high abundance in bodily fluids. Examples include albumin fragments as biomarkers for renal disease found by CE-FT-ICR MS in urine (Weissinger *et al*., 2004); α-2 macroglobulin and transthyretin which are differentially expressed in plasma samples of AD and normal individuals, found by 2D electrophoresis coupled with tandem mass spectrometry (Zhang *et al*., 2004a); and the C3 and C4 components of the complement pathway of the immune system and C1 inhibitors, which differed between plasma of healthy individuals and those with type 2 diabetes (Zhang *et al*., 2004a). The findings of such studies require confirmation in large series by ELISA (Allard *et al*., 2004) or by genetic studies of the candidate biomarkers. Furthermore, proteins of the immune system found differentially expressed between disease and healthy samples need further verification as to whether they are merely a non-specific result of inflammation.

Secondly, high levels of variation observed between replicate samples in the analysis of proteins by MALDI-TOF MS and CE could be the result of cumulative variation at different stages of sample pre-treatment. Furthermore, factors such as sample freeze-thaw cycles, clotting time and storage temperatures have been shown to affect variability between experiments and across different study groups (de Noo *et al*., 2005; Baumann *et al*., 2005).

Thirdly, instrumental limitations could explain the high level of variation observed between replicates in MALDI-TOF MS and lack of detection of proteins by CE FT-ICR MS. For example, the lack of consistent detection of protein species larger than 100 kDa and smaller than 10 kDa, due to the failure to ionise or move efficiently through the time-of-flight chamber, was a major limitation of the MALDI-TOF MS technique. Failed attempts at protein detection by CE coupled with FT-ICR MS were also due to technical problems associated with the instruments involved that could not be explained.

In conclusion, despite the lack of success in optimising MALDI-TOF MS and CE FT-ICR MS techniques for protein detection and quantitation, the findings of the IEC coupled with nLC-MS/MS in relation to quantitation of the full length plasma CFH were promising. This was achieved by using plasma CFH measurements

obtained by ELISA as internal controls to which IEC results were normalised. Whether this technique would be successful in identification and quantitation of novel biomarkers in human plasma requires further research.

CHAPTER 9

DISCUSSION

Background

The aims of this study were to analyse known and potential biomarkers of common and genetically complex human disorders, and to identify genetic and environmental variation associated with plasma biomarker concentrations. The aims of the project were drawn from the following hypotheses. Firstly, based on the genetic architecture of complex traits, intermediate phenotypes have stronger genetic determinants than the downstream phenotype they mediate (Falchi *et al*., 2004). Secondly, genetically simplified isolated populations, with a more uniform distribution of environmental risks and reduced genetic heterogeneity provide optimal conditions for mapping complex disorders (Wright *et al*., 1999).

Two groups of protein biomarkers were analysed. First, plasma complement factor H (CFH) was selected as a potential biomarker for age-related macular degeneration (AMD), since common coding and non-coding variants in the *CFH* gene have been reported to show strong association with this disorder (Klein *et al*., 2005; Haines *et al*., 2005; Edwards *et al*., 2005; Zareparsi *et al*., 2005a). Secondly, two isoforms of amyloid-β (Aβ40 and Aβ42) were selected as biomarkers for Alzheimer disease (AD), since Aβ deposits in the brain are major constituents of the amyloid plaques characteristic of this disorder (Glenner & Wong, 1984; Masters *et al*., 1985). A third approach was to use mass spectrometric techniques to identify new biomarkers for disease by analysis of plasma protein samples from an isolate population.

Experimental overview

Completion of the human genome sequencing and the advent of high throughput genotyping technologies have allowed large-scale genetic analysis of polymorphisms which influence susceptibility to common and complex human disorders.

Physiological and anthropometric measurements and samples of human plasma and genomic DNA were collected from a population sample of 1,021 individuals from the isolated Croatian island of Vis after an initial epidemiological survey of the island (Rudan *et al*., 1999b). Extensive genealogical data were also obtained from the island which later allowed construction of pedigrees within the population sample.

Firstly, the availability of human plasma samples allowed the biochemical analysis of known (Aβ40 and Aβ42) and potential (CFH) biomarkers of disease using enzyme-linked immunosorbent assays. Attempts were also made to identify and quantitate new biomarkers in human plasma, using techniques such as MALDI-TOF mass spectrometry, capillary electrophoresis and ion exchange chromatography. Secondly, genome-wide linkage and association analyses were conducted after an initial assessment of power, in order to identify novel quantitative trait loci (QTL) which influence plasma concentrations of CFH, Aβ40 and Aβ42.

Linkage versus association analysis

Linkage and association analyses are two of the most powerful genome-wide approaches for the identification of variants underlying complex traits, in which locus heterogeneity, epistasis, low penetrance and pleiotropy create special challenges (Lander & Schork, 1994). Linkage studies in humans have been successful in mapping susceptibility variants of moderate to large effect, such as *APOE*E4* as a strong and common predisposition locus in Alzheimer disease (Pericak-Vance *et al*., 1991; Corder *et al*., 1993). Similarly, the MHC genes, *HLA-DR* and *HLA-DQ*, have been identified as relatively strong susceptibility loci for type 1 diabetes (Todd *et al*., 1987). Genome-wide association studies are more recent phenomena but have already achieved great success in identifying replicable common variants of moderate to large effect, such as the *CFH* Y402H polymorphism predisposing to age-related macular degeneration (Klein *et al*., 2005). The identification of novel associations with bipolar disorder, coronary artery disease, Crohn's disease, rheumatoid arthritis, type 1 and type 2 diabetes using large case-control studies (2,000 cases and 3,000 controls) was also a major success (Wellcome Trust Case Control Consortium, 2007).

In the present study, genome-wide quantitative trait linkage analysis of plasma Aβ and CFH was unsuccessful in detecting loci with either suggestive (LOD > 2) or significant (LOD > 3) linkage as defined previously (Lander & Kruglyak, 1995). The lack of success of the linkage analysis can be due either to the small size of the study population (especially the size of the pedigrees) or to the lack of power of linkage to detect variants of modest effect size (Dawn & Barrett, 2005). In the

case of plasma Aβ40 and Aβ42, the failure of linkage and the success of association could be explained by the lack of significant narrow-sense heritabilities (h^2) observed for those traits. The power in genome-wide analyses has been shown to be directly related to the QTL heritability in the case of association but to the square of the QTL heritability in the case of linkage (Sham *et al*., 2000). Therefore, with decreasing QTL heritability, association will become progressively more powerful than linkage. However, in the case of plasma CFH, where a relatively high heritability of 46% was observed and the underlying QTL was found by association to be of large effect size (explaining 28% of the trait variance) the lack of power in linkage can be explained by the small size of the pedigrees. Using simulations, Sham and colleagues (2000) estimated that approximately 20,000 sib-pairs would be required to detect a QTL of 10% effect size (i.e. explaining 10% of the QT variance) using linkage with 80% power. The number of required sib-pairs decreases to approximately 1,000 for detection of statistically significant associations, assuming that the degree of linkage disequilibrium between the causal QTL and the marker is strong ($r^2 > 0.50$).

One way of assuring maximum power is achieved in linkage is to use simulations as part of the study design, which takes the particular features of the study, including family structure, marker map, allele frequencies and patterns of missing data into account, according to the null hypothesis of no linkage across the whole genome (Dawn & Barrett, 2005). This method has been applied to genome-wide linkage studies of plasma HDL (Arya *et al*., 2002), myocardial infarction (Broeckel *et al*., 2002), hypertension (Caulfield *et al*., 2003) and schizophrenia (Williams *et al*., 2003) in which the correct significance level has been determined for the specific study.

Quality control prior to the association analyses

In this study, markers with more than 5% missing genotypes were excluded as part of the quality control procedure before the association analyses. In general SNPs are of questionable quality if their genotypes failed in many individuals or if the calling algorithm was not able to assign a genotype for many subjects. Thresholds of 3% (Wellcome Trust Case Control Consortium, 2007) and 2% (Samani *et al*., 2007) have been used previously in the literature. This criterion should be set separately for

different study groups as the DNA quality and the genotyping platforms may differ between studies.

Similarly, individuals with more than 10% missing genotypes (i.e. > 90% call rate) were excluded in this study as part of a subject-wide quality control measure. Higher thresholds of subject-wide genotype call rate may be preferred. However, subjects should initially be excluded based on a less stringent threshold of 10% missing call rate. Genotype call rates can consequently be checked for each of the trait-associated markers and "re-called" based on a higher threshold. Ambiguous genotypes can then be excluded. This method will reduce the probability of reporting false-positive and -negative associations. The heterozygosity across all SNPs can also be a good quality indicator, such that high heterozygosity (mean ± 3 SD) may indicate DNA sample cross-contamination (Ziegler *et al*., 2008).

In an assessment of population substructure in the Vis sample set using the Illumina Hap300 genotype data, after exclusion of outliers, subjects were classified into two clusters corresponding approximately to the two villages of Komiza and Vis based on excess allele sharing. Adjustment for population stratification within the two clusters defined by allele sharing was performed using a permutation approach implemented in the whole-genome data analysis tool set PLINK (Purcell *et al*., 2007). Population substructure can generate spurious genotype-phenotype associations as the result of genetic markers with allele proportions which differ between the subgroup and the general population (Balding, 2006). The statistical method of genomic control, which uses markers throughout the genome to adjust for any inflation in test statistics due to substructure, has been used to overcome spurious associations induced by population substructure in case-control studies (Devlin & Roeder, 1999).

Quality control post-association analyses

In order to control for multiple testing in single SNP association analyses, the significance level, α ($\leq 5\%$), was set so that all the tests together generate no more than the probability of a false positive, α. The most conservative application of this theory is the Bonferroni correction, $\alpha' \approx \alpha / n$, assuming a number n of approximately independent tests (Miller, 1981). However, the Bonferroni correction

is conservative in that it does not take account of tightly linked or associated loci. Alternatively, permutation procedures and false discovery rate (FDR) estimation have been employed to correct for multiple testing. Permutation procedures involve retaining the genotype data (i.e. the observed LD structure) but randomising phenotype labels over individuals (Balding, 2006). Analysing many such datasets can approximate the false positive rate. However, the disadvantage of this method is that it is computationally demanding. The other approach is to monitor the FDR, which is the proportion of false positive test results among all positives (Storey & Tibshirani, 2003). In other words, the distribution of *P*-values is estimated as a mixture of outcomes under the null (uniform distribution of *P*-values) and alternative hypotheses (*P*-value distribution skewed towards zero). The disadvantage of this method is that an expectation of the number of false positives is required prior to the analysis.

In the association analysis of plasma CFH, the conservative Bonferroni correction of multiple testing was used since a large number of markers showing highly significant genome-wide association ($P \leq 1.6E-07$) were detected at the *CFH* locus. In the association analysis of plasma Aβ, no markers passed the threshold of significance after Bonferroni correction for multiple testing. In this analysis, less stringent thresholds of genome-wide significance were therefore set in return for greater power. Consequently, associations worthy of further investigation were retained for replication or further analysis.

All single-marker associations were checked for deviations from Hardy-Weinberg equilibrium using a Fisher's exact test (Guo & Thompson, 1992; Wigginton *et al*., 2005). Testing for HWE deviations can also be carried out using a Pearson goodness-of-fit test, which has approximately a χ^2 null distribution but can be unreliable when there are low genotype counts. *CFH* Y402H was the only Illumina SNP which showed deviation from HWE in the Vis dataset. This was investigated and found to be due to genotyping errors. Consequently genotyping of this SNP was carried out using TaqMan and the genotype data were replaced. Deviation from HWE can be the result of faulty SNP genotyping. Deviations from HWE can also be due to inbreeding, population stratification or selection (Balding, 2006). Deviations can also arise in the presence of a common deletion

polymorphism, because of a mutant PCR primer site or because of a tendency to miscall heterozygotes as homozygotes. The SNP rs1066420 was found to significantly deviate from HWE in both Vis and Scottish AMD data sets and was consequently removed from the analysis. However, this SNP is believed to be located within the polymorphic deletion of *CFHR3/CFHR1*, resulting in deviation from HWE. One way of detecting false genotype data is to check for Mendelian errors of inheritance. This was carried out for all markers prior to the genome-wide association analysis and polymorphisms showing errors in patterns of inheritance were excluded.

Deviations from HWE have so far been mainly tested as a measure of data quality control. However, since deviations from HWE can be due to deletion polymorphisms (Conrad *et al*., 2006) or segmental duplications (Bailey & Eichler, 2006), test of HWE was used as a quality control check after the association analysis in order to prevent exclusion of markers with potentially interesting trait associations as previously recommended (Zou & Donner, 2006).

Analysis of plasma amyloid-β

Quantitative determination of Aβ40 and Aβ42 in CSF, brain tissue and cell lines by ELISA have been reported in the literature (Gravina *et al*., 1995; Kuo *et al*., 1998; Lewczuk *et al*., 2004; Bibl *et al*., 2004; Duering *et al*., 2005). However, there have been conflicting results in the literature about the correlation between concentrations of Aβ in CSF and plasma (Mehta *et al*., 2001; Fukumoto *et al*., 2003). The problem arises when trying to determine concentrations of Aβ in plasma where they are about 50-fold lower than those observed in CSF or brain. Furthermore, large quantities of Aβ have been shown to be sequestered by plasma proteins, in particular serum albumin, potentially preventing them from being detected by immunoassays (Kuo *et al*., 1999). This interaction is thought to be the result of the high reactivity of Aβ due to its amphipathic and amphoteric structure. The C-terminal region of Aβ is composed of 12-14 hydrophobic amino acids derived from the APP transmembrane domain. Conversely, the N-terminal 28 amino acids are mainly hydrophilic, with the exception of a stretch of five non-polar side chains localised at residues 17-21. The

N-terminal of Aβ is also highly charged at physiological pH, carrying six negative and six positive charges, which are believed to aid its binding to carrier molecules. Albumin, the most abundant component of human brain homogenates, not only acts as a carrier protein but also interferes with the processing of Aβ-containing peptides *in vivo* by masking Aβ epitopes from proteases, inhibitors and matrix proteins (Matsumoto *et al*., 1997). Another study showed that large proteins in plasma, such as α-2 macroglobulin and transthyretin, sequestered Aβ peptides *in vitro* (Kuo *et al*., 2000). These authors found that greater than 95% of Aβ40 and Aβ42 in plasma resides within a sequestered pool bound to proteins such as albumin. The development of robust assays, which overcome the problem of interference and sequestration by high abundance plasma proteins, would be beneficial in reliable quantitation of plasma Aβ concentrations. The antibodies used in the current study for the measurements of Aβ in EDTA plasma had a higher affinity for Aβ species than serum albumin and hence resulted in improved sensitivity in the Aβ assay compared with the commercial assay (Suzuki *et al*., 1994). However, development of techniques to release the bound fraction of Aβ in plasma would also be more advantageous for the quantitation of plasma Aβ.

APOE genotypes are not associated with plasma Aβ

APOE has been consistently shown to influence genetic predisposition to Alzheimer disease (Corder *et al*., 1994; Rubinsztein & Easton, 1999). We hypothesised that APOE isoforms may regulate concentrations of circulating Aβ species in plasma and thus influence susceptibility to AD. This hypothesis was tested genetically, in univariate analyses of *APOE* allele dosage and plasma Aβ concentration, which showed no significant association between plasma Aβ and *APOE*. Despite the lack of genetic evidence for direct interaction between APOE isoforms and Aβ in plasma, the interaction may be localised to the brain or at the brain-blood barrier where APOE has been shown to be highly localised. The APOE protein is produced in abundance in the brain and induced at high concentrations in peripheral nerve injury. It is plausible that the APOE isoforms differentially affect amyloid plaque formation by direct protein-protein interaction, Aβ peptide metabolism, or both. Lipid-free

APOE*E4 has been shown to form a stable complex with Aβ peptide *in vitro* (Stratman *et al*., 2005). In contrast, lipidated APOE*E3 has been found to bind Aβ peptide with a 20-fold higher affinity than lipidated APOE*E4, which may prevent the accumulation of the neurotoxic Aβ species (Stratman *et al*., 2005; Mahley & Huang, 1999). It has also been suggested that the extra-cellular accumulation of Aβ is due in part to a direct physico-chemical interaction with cholesterol in neuronal plasma membranes (Wood *et al*., 2002). *In vitro* interaction between APOE isoforms and Aβ40 has been shown in hamster kidney cells (Aleshkov *et al*., 1997) and *APOE2* knock-in mice have been reported to be protected against AD (Mann *et al*., 2004), indicating possible involvement of APOE in the binding and clearance of Aβ in a tissue-specific manner.

Genetic analysis of plasma Aβ

Both species of plasma Aβ (Aβ40 and Aβ42) have been used to map genomic loci involved in Alzheimer disease, using series of late-onset AD cases and controls (Ertekin-Taner *et al*., 2000; Ertekin-Taner *et al*., 2005). Two genes, *IDE* (Ertekin-Taner *et al*., 2000; Prince *et al*., 2003; Ertekin-Taner *et al*., 2004) and *PLAU* (Ertekin-Taner *et al*., 2005), on chromosome 10q were identified in independent cohorts, using linkage and association approaches. However, in these studies families had been ascertained through AD probands presenting with high concentrations of plasma Aβ, in particular Aβ42. To date no genome-wide association studies of plasma Aβ, in samples of unselected individuals have been reported. In this study, genomic loci with potential relevance to Alzheimer disease were identified in a general population sample using a genome-wide association analysis approach. None of the associated SNPs were significant after correction for multiple testing, so that replication is required. However, some of the loci showing suggestive association with plasma Aβ have roles in neuronal migration and inflammatory response with high reported expression levels in the brain and therefore are good candidates for influencing plasma Aβ concentrations.

There are numerous reports of positive associations in late-onset AD which have failed in replication studies (Prince *et al*., 2001). Lack of success in replication

is dependent upon the size of the effect of a particular polymorphism and the size of the population studied. The replication of the Aβ results in an independent data set has become feasible since the Croatian field work has now been extended to include the neighbouring isolated island of Korcula (Sujoldzić *et al*., 1989; Waddle *et al*., 1998; Rudan *et al*., 1999b; Klarić *et al*., 2001) where a further 1,000 individuals have been sampled and are currently undergoing a genome-wide association scan. Also, collaboration within the European Special Populations Research Network (EUROSPAN) consortium of isolate populations in Sweden, Netherlands, Croatia, Scotland and Italy (led by Professor H. Campbell, University of Edinburgh) will allow for replication and comparison of results in different isolate and urban populations.

Putative associations detected in the current study could initially be replicated in the Korcula sample set and further validated in large series of Alzheimer disease cases and controls available at the Mayo Clinic in Florida. Polymorphisms with confirmed roles in influencing plasma CFH concentrations can be analysed further by assessing the biochemical pathways in which they are involved. Functional studies of potential regulatory loci can be carried out *in vitro* using luciferase reporter assays (Kanda *et al*., 2007) and *in vivo* using transgenic animal models or by injection of reporter constructs into zebrafish embryos (Patton & Zon, 2005; Kleinjan *et al*., 2008). Assessment of protein interactions using techniques such as yeast-two-hybrid can also reveal interaction partners for the associated proteins.

Analysis of CFH in the context of AMD

The full length human CFH is a multidomain, multifunctional plasma glycoprotein which acts as a cofactor for the factor I-mediated degradation of C3b and accelerates the decay of C3 convertase, down-regulating the alternative complement pathway (Whaley & Ruddy, 1976). CFH is a member of a group of structurally and immunologically related plasma proteins which consist of individually folding protein domains called short consensus repeats (SCRs) (Zipfel & Skerka, 1994). The CFH protein also binds cellular receptors such as CD18 (DiScipio *et al*., 1998), heparin (Meri & Pangburn, 1990), the acute phase protein C-reactive protein (CRP), although there is now some doubt about this (Jarva *et al*., 1999) and the surface of

certain pathogenic microorganisms such as *S. pyogenes*, *N. gonorrhoeae* and *B. afzelii* (Lindahl *et al*., 2000).

Plasma CFH deficiency resulting from *CFH* truncating mutations have been linked to rare, severe kidney defects in membranoproliferative glomerulonephritis type II (MPGN II) (Levy *et al*., 1986; lopezlarrea *et al*., 1987) and atypical haemolytic uraemic syndrome (aHUS) (Warwicker *et al*., 1998; Richards *et al*., 2001; Caprioli *et al*., 2003). Genetic variants in the *CFH* genomic region have also shown strong associations with age-related macular degeneration (AMD) (Haines *et al*., 2005; Edwards *et al*., 2005; Klein *et al.*, 2005, Zareparsi *et al*., 2005a; Hughes *et al*., 2006; Li *et al*., 2006).

This study aimed to identify genetic and environmental factors which influence plasma CFH concentration and hence affect predisposition to AMD and perhaps other CFH-associated disorders such as MPGN II or aHUS. Plasma CFH concentration was measured using an optimised ELISA and covariate analyses showed that it is significantly influenced by sex, age-by-sex interaction and BMI. Narrow-sense heritability of plasma CFH was also estimated to be 0.46, accounting for the additive genetic effect of variants throughout the genome. The heritability of plasma CFH has previously been reported to be 0.63 in an urban Spanish population (Esparza-Gordillo *et al*., 2004).

Genome-wide linkage analysis of plasma CFH in the Vis series failed to identify regions of suggestive ($Z > 2$) or significant ($Z > 3$) linkage despite the fact that genome-wide association analysis of plasma CFH revealed genetic polymorphisms of large effect influencing plasma CFH concentration. The lack of power in linkage compared with association in the Vis samples could be the result of low power to detect QTL with low QTL-specific heritabilities. Analysis of power to detect QTL using the variance components methodology, given the pedigree structure in Vis, has shown that for a fixed trait heritability of 0.40, there would only be sufficient power (> 0.50, $\alpha = 0.05$) to detect a QTL-specific heritability of 0.30 (Dr. V. Vitart, personal communication). This indicates that there is insufficient power in the Vis sample set to detect QTL with heritabilities smaller than 0.30, probably due to the family structure and the dependency of linkage on the QTL-specific heritability. However, the predictive analysis of power, showed that there is

sufficient power to detect variants of effect size as low as 3% using association in 500 unrelated individuals from Vis assuming that the marker and the QTL are in strong linkage disequilibrium.

The SNP most strongly associated with plasma CFH, rs6677604, is a non-coding polymorphism in *CFH* intron 11 which was found to explain 28.3% of the covariate adjusted phenotypic variance in the Vis population and 21.8% in the Dutch. Furthermore, the *CFHR3/CFHR1* deletion polymorphism was found to be in strong linkage disequilibrium with *CFH* rs6677604 (Hughes *et al*., 2006; Hageman *et al*., 2006). The results of the genome-wide association analysis showed that the *CFH* Y402H variant which has been strongly implicated in predisposition to AMD (Haines *et al*., 2005; Edwards *et al*., 2005; Klein *et al.*, 2005, Zareparsi *et al*., 2005a) also influences plasma CFH concentration although the effect is weak compared with *CFH* rs6677604.

The genomic region showing association with plasma CFH covered a region of approximately 300 kb from *CFH* to *CFHR2*, which consisted of many polymorphisms falling into several LD groups. In order to identify a multi-marker combination of the SNPs which might explain more of the variation in plasma CFH than SNP rs6677604, a 'backward' regression analysis was carried out. This resulted in a 3-SNP model, containing *CFH* SNPs rs6677604 (*CFH* intron 11), rs1292487 (1.6 kb upstream of *CFH*) and rs10801575 (4.4 kb upstream of *CFHR4*), with age and BMI as covariates. The 3-SNP haplotype which covered a region of approximately 230 kb was associated with an increase of 55.5 μg/ml in plasma CFH concentration compared with 59.4 μg/ml associated with the SNP rs6677604 alone ($P > 0.05$). Analysis of complex traits in the context of haplotypes can be advantageous because protein products of the candidate genes occur in polypeptide chains whose folding and other properties may depend on particular combinations of amino acids (Clark, 2004). Furthermore, in any given population, factors such as mutation, drift, selection, migration and recombination result in genetic variation that has a strong segment-wise haplotype structure to it (Clark, 2004).

The 3-SNP model was not shown to have a larger effect on plasma CFH than the single SNP, rs6677604, but it could help to understand the mechanism by which SNP rs6677604 may be influencing plasma CFH concentration. For example,

including the SNPs rs6677604 and rs1292487 in the 3-SNP haplotype fitted with the DNA looping model during transcription (**Figure 9.1**) and the interaction between intronic enhancer elements with the RNA polymerase machinery located in the promoter region of genes during transcription (Matthews, 1992; Nussinoc, 1992; Ogata *et al*., 2003).

The role of the *CFH*-related proteins in regulation of complement is unknown. However, SNP rs10801575 and the 3′ end of the *CFH* gene are in close proximity on chromosomes containing the *CFHR3/CFHR1* deletion in which case the linkage disequilibrium between the SNP rs10801575 and polymorphisms in *CFH* could explain the effect associated with this SNP.

Haplotype analysis conditional on the *CFH* rs1061170 (Y402H) or the *CFHR3/CFHR1* deletion polymorphism showed that the effect associated with the haplotype was independent of both rs101160 (*CFH* Y402H) and the deletion polymorphism.

In a case-control analysis of plasma CFH concentrations in the Scottish AMD series significant differences were observed between cases of AMD compared with disease-free controls in plasma CFH, which was found to explain 1.8% of the AMD variance. This has potentially important implications for treatment of AMD since delivery of additional CFH using therapy to the RPE might contribute to disease prevention in high-risk individuals.

Case-control analysis of AMD susceptibility loci showed that the SNPs identified to be strongly associated with plasma CFH concentration influenced AMD susceptibility independently of the *CFH* Y402H polymorphism. The strongest genetic risk factor for AMD in the Scottish series was *CFH* SNP rs1061170 (Y402H). In contrast, the effect of the risk haplotype containing *CFH* rs2274700 (Li *et al*., 2006) was non-significant when conditioned on the *CFH* Y402H polymorphism implying non-independence. Furthermore, the *CFH* SNPs rs2274700 and rs6677604 together with the *CFHR3/CFHR1* deletion polymorphism were found to collectively affect AMD susceptibility independently of rs1061170 (Y402H) using conditional haplotype analyses, consistent with their effects on plasma CFH concentration.

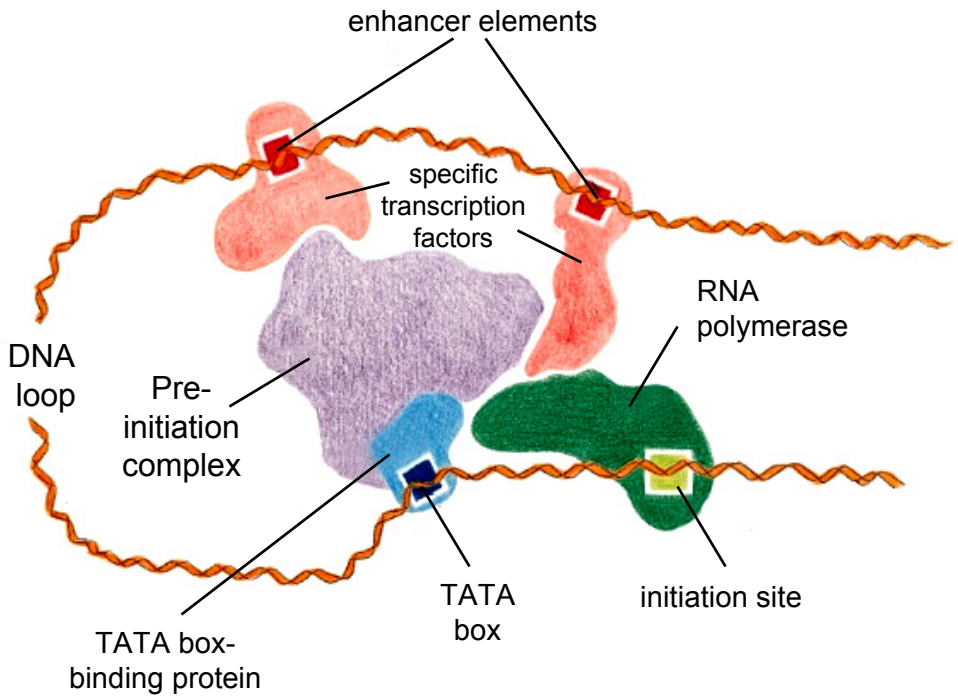


Figure 9.1 A simplified diagrammatic presentation of a eukaryotic gene during transcription.

Diagram shows the DNA looping mechanism during transcription by the RNA polymerase complex. The diagram demonstrates the interaction between intronic regulatory (enhancer) elements and transcription factor subunits which are bound directly to the RNA polymerase itself or the transcription pre-initiation complex located at the transcription start site. Image reproduced from http://pps00.cryst.bbk.ac.uk/course/section10/dna.html.

Based on these results, a model is proposed for AMD susceptibility which incorporates *CFH* rs1061170 (Y402H), *ARMS2* rs10490924 and plasma CFH concentration as significant risk factors for AMD (**Figure 9.2**). The proposed model for AMD shows the *CFH* SNP rs1061170 (Y402H) and *ARMS2* SNP rs10490924 as the strongest genetic susceptibility loci for AMD explaining 6.8% and 6.4% of the disease variance respectively. The rs1061170 polymorphism which occurs in CFH module SCR7, binds a heparin-like glycosaminoglycan (GAG) (Prosser *et al*., 2007) and strong evidence from crystal structure analysis shows that GAG binding is altered by this amino acid substitution (Prosser *et al.*, 2007). This supports the case for a causal link between the polymorphism and a mechanism for AMD involving insufficient complement regulation in the ageing choroid. The proposed model for AMD shown in **Figure 9.2** shows that *CFH* SNPs rs6677604, rs2274700 and the *CFHR3/CFHR1* deletion polymorphism significantly affect AMD susceptibility by influencing the plasma CFH concentration, accounting for 1.8% of the AMD disease variance. This effect is found to be independent of the *CFH* rs1061170 (Y402H) which also shows a weaker association with changes in plasma CFH concentration but is postulated to have a strong effect on disease via functional effects on GAG binding (Prosser *et al*., 2007). The combination of these effects in rs1061170 may explain its large effect on disease susceptibility.

Identification of the *CFH* Y402H polymorphism as the strongest associated genetic risk factor for AMD contradicts some published work, which identified the *CFH* SNP rs2274700 (Ala473Ala) as the single strongest genetic susceptibility factor for AMD (Li *et al*., 2006; Hughes *et al*., 2006). Differences observed in the relative strength of association for rs1061170 and rs2274700 with AMD in different populations could be due to genetic differences between geographic sub-groups of individuals, irrespective of their disease status (population stratification) (WTCCC, 2007). The WTCCC authors reported allele frequency differences between samples of individuals from 12 geographical regions across Great Britain which have resulted in false-positive association findings in 16 genomic regions.

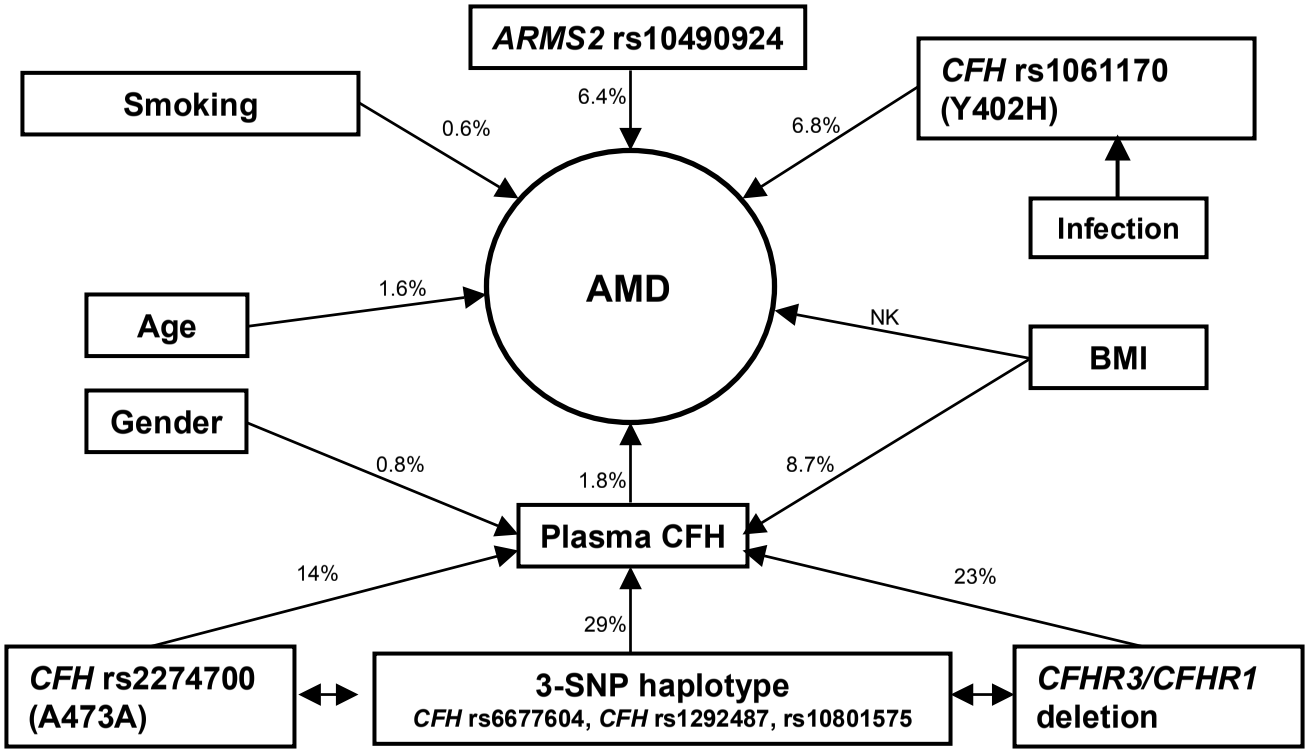


Figure 9.2 A proposed model for the relationship between AMD susceptibility factors including those affecting plasma CFH concentration based on the findings of the present study.

Diagrammatic presentation of the complex structure and relationship between AMD susceptibility factors based on data from the Scottish AMD case-control cohort and the Vis isolate. Arrows represent the direction of the effect. The percentages of the phenotypic variance explained by each factor are displayed next to each arrow. Double-headed arrows indicate interactions. The association between AMD and BMI has been taken from published data (Gold *et al*., 2006) for which the the proportion of disease variance explained is not known (NK). The association between *C3* and AMD has also been shown in this cohort and described elsewhere (Yates *et al*., 2007).

Although the 1q31 chromosomal region was not among the regions showing geographic differences in the UK population, the possibility of sample heterogeneity in the U.S. population in relation to the association between *CFH* rs2274700 and AMD due to the presence of genetically stratified sub-groups cannot be excluded.

A haplotype which contains the *CFH* Y402H risk allele has been shown to be the most divergent (compared with haplotypes containing the *CFHR3/CFHR1* deletion polymorphism) and least related to the sequence found in the chimpanzee, perhaps reflecting recent selection pressure acting on this allele (Hageman *et al*., 2006). The *CFH* H^{402} variant has been shown to be less tightly sequestered than the *CFH* Y^{402} variant by the M6 protein of *S. pyogenes* and might confer an evolutionary advantage in that respect (Yu *et al*., 2007). Furthermore, there is now evidence that the RCA cluster containing the *CFH* and *CFH*-related genes has been under strong selection pressure, based on information from common SNPs (MAF > 0.05) available in the HapMap database (Voight *et al*., 2006).

Reports have shown that the *CFHR3/CFHR1* deletion polymorphism is included in an extended protective haplotype which contains *CFH* SNPs rs6677604 and rs2274700 (Hughes *et al*., 2006; Hageman *et a*l., 2007). This association is confirmed in this study and the protective effect associated with the haplotype containing the *CFH* rs6677604 SNP is shown to be independent of *CFH* rs1061170 (Y402H) and more significant than *CFH* rs2274700 (Ala473Ala) or the *CFHR3/CFHR1* deletion. A hypothesis that explains this result is that the regulatory polymorphisms in the non-coding sequence of *CFH* which affect *CFH* expression, are influencing plasma CFH concentration and hence affecting predisposition to AMD. Furthermore, the reported associations between AMD and the *CFH* SNP rs2274700 (Li *et al*., 2006) and *CFHR3/CFHR1* deletion polymorphism (Hughes *et al*., 2006; Hageman *et al*., 2007) can be explained by a regulatory effect such as an enhancer sequence either close to rs6677604 or within the 3-SNP haplotype identified in this study. However, the 3-SNP haplotype only narrowed the associated region to 355 kb, in which sequence analysis identified many areas of "regulatory potential", based on the alignments of human, chimpanzee, macaque, mouse, rat, dog and cow genomic sequences (Kolbe *et al*., 2004; King *et al*., 2005). Therefore, as a starting point a 5 kb region spanning the highest associated peak, rs6677604, has

been selected for further analysis. This region was shown to contain several stretches of DNA showing "regulatory potential" which are highly conserved in mammals and fish and would therefore make a plausible candidate enhancer region for regulation of gene expression in *CFH*.

Two approaches are being tested to assess whether a construct containing the 5 kb region would drive expression of a luciferase reporter gene containing a basal promoter. Firstly, cellular-based luciferase assays involving liver (HepG2) and RPE (ARPE-19) cell lines will be used to assess whether the construct can drive expression of the luciferase reporter higher than the basal promoter construct (Kanda *et al*., 2007). Secondly, *in vivo* analysis of the same region using a different plasmic construct is possible by using the zebrafish embryo as a model (Kleinjan *et al*., 2008). This method involves injection of a construct containing the 5 kb region and a GFP/YFP reporter, into zebrafish oocytes and measuring the expression of the reporter compared with the background noise. Despite being separated by about 100 million years, there is a high level of conservation in the *CFH* genomic region between human and fish (Kemper *et al*., 1998, 2000). Expression of the putative regulatory region in zebrafish embryos is expected to be largely concentrated in the liver as the main site of *CFH* expression and secretion.

Mass spectrometry and biomarker analysis

A number of mass spectrometry techniques were assessed for their ability to identify and/or quantitate variation in human protein concentrations in plasma samples from the general population of Vis. Two main approaches of "top-down" (including MALDI-TOF MS and CE FT-ICR MS) and "bottom-up" (IEC nLC-MS/MS) were employed (Chait, 2006).

The two top-down approaches were largely unsuccessful in reliably identifying proteins in complex mixtures. Quantitation was also not possible due either to variability or lack of signal. Ion exchange chromatography of the digested plasma proteins, after separation by electrophoresis resulted in more promising results. The lack of success in mass spectrometric techniques in protein quantitation is mainly due to the complexity of the fractions remaining even after immunodepletion or fractionation procedures. If the complexity of the eluate

reaching the MS instrument is high, low concentrated proteins or peptides remain undetected even if the sensitivity limit of the instrument has not yet been reached. This suppression effect results from the competition of too many analyte molecules for protons during the ionisation process (Frohlich & Arnold, 2006).

Mass spectrometry techniques such as iTRAQ, involving labelling of digested proteins have been suggested to be the only method of reliably assessing protein concentrations (relative to an internal standard) in complex mixtures (Ross *et al*., 2004; DeSouza *et al*., 2005). However, currently the major disadvantage of using labelling techniques is the cost of the reagents involved. The lack of expertise in the analysis of results using complex pattern recognition software is a second area in need of improvement (Shadforth *et al*., 2005).

Summary

Three approaches for biomarker discovery and quantitation were employed in this study. Firstly, plasma CFH was chosen as a surrogate biomarker for age-related macular degeneration since genetic variation in the *CFH* gene has been shown to affect susceptibility to AMD. This study resulted in the identification of plasma CFH as a biomarker for AMD and the discovery of several polymorphisms including *CFH* SNP rs1061170 (Y402H) with highly significant effects on plasma CFH.

Secondly, plasma Aβ species were used as biomarkers for Alzheimer disease in order to identify novel loci which influence susceptibility to AD. Measurements of plasma Aβ in general population samples resulted in the identification of several potentially relevant genomic loci influencing plasma Aβ concentrations. The effects associated with the putative loci in the context of Alzheimer disease will need to be assessed after replication of the association findings in independent samples.

Thirdly, proteomic approaches were employed to discover new biomarkers for disease in plasma samples collected from Vis and to assess their importance by analysing the protein profiles and quantitating their concentrations. The label-free techniques, MALDI-TOF and CE FT-ICR mass spectrometry were largely unsuccessful due to interference from high abundance proteins and instrument limitations. More promising results were obtained from another label-free technique, ion exchange chromatography coupled with tandem mass spectrometry. Assessment

of the iTRAQ method for protein quantitation was not possible due to shortage of time and the high cost of the reagents.

Finally this study demonstrated the advantage of using biomarkers of disease to identify genetic and environmental risk factors. This approach should be applied to other newly identified genetic risk factors of AMD such as plasma complement component 3 (C3) (Yates *et al*., 2007), complement factor B (CFB) and complement component 2 (C2) (Gold *et al*., 2006) since it is highly likely that genetic variants affecting plasma concentrations of these proteins influence the disease. It should also be noted that although biomarkers can explain a significant proportion of a complex disease, the major disease susceptibility loci may be acting through some other unknown mechanism in which case additional approaches are required, such as case-control association studies. The strong associations between *APOE* and Alzheimer disease despite the lack of significant association of *APOE* with Aβ clearly illustrates this point.

By using biomarkers of disease, which have a simplified genetic architecture, this study identified a major *CFH* regulatory locus affecting age-related macular degeneration, as well as several loci with potential roles in influencing susceptibility to Alzheimer disease. These findings will help to unravel part of the complex genetic architecture of two of the most common and complex human disorders.

CHAPTER 10
REFERENCES

Abney, M., McPeek, M. S., & Ober, C. Broad and narrow heritabilities of quantitative traits in a founder population. *Am J Hum.Genet.* **68**, 1302-1307 (2001).

Adkins, J. N., Monroe, M. E., Auberry, K. J., Shen, Y. F., Jacobs, J. M., Camp, D. G., Vitzthum, F., Rodland, K. D., Zangar, R. C., Smith, R. D., & Pounds, J. G. A proteomic study of the HUPO Plasma Proteome Project's pilot samples using an accurate mass and time tag strategy. *Proteomics* **5**, 3454-3466 (2005).

Adkins, J. N., Varnum, S. M., Auberry, K. J., Moore, R. J., Angell, N. H., Smith, R. D., Springer, D. L., & Pounds, J. G. Toward a human blood serum proteome - Analysis by multidimensional separation coupled with mass spectrometry. *Molecular & Cellular Proteomics* **1**, 947-955 (2002).

Aggarwal, K., Choe, L. H., & Lee, K. H. Quantitative analysis of protein expression using amine-specific isobaric tags in Escherichia coli cells expressing rhsA elements. *Proteomics* **5**, 2297-2308 (2005).

Aleshkov, S., Abraham, C. R., & Zannis, V. I. Interaction of nascent ApoE2, ApoE3, and ApoE4 isoforms expressed in mammalian cells with amyloid peptide beta (1-40). Relevance to Alzheimer's disease. *Biochemistry* **36**, 10571-10580 (1997).

Allard, L., Lescuyer, P., Burgess, J., Leung, K. Y., Ward, M., Walter, N., Burkhard, P. R., Corthals, G., Hochstrasser, D. F., & Sanchez, J. C. ApoC-I and ApoC-III as potential plasmatic markers to distinguish between ischemic and hemorrhagic stroke. *Proteomics* **4**, 2242-2251 (2004).

Allikmets, R. Further evidence for an association of ABCR alleles with age-related macular degeneration. The International ABCR Screening Consortium. *Am J Hum.Genet.* **67**, 487-491 (2000).

Allikmets, R., Shroyer, N. F., Singh, N., Seddon, J. M., Lewis, R. A., Bernstein, P. S., Peiffer, A., Zabriskie, N. A., Li, Y., Hutchinson, A., Dean, M., Lupski, J. R., & Leppert, M. Mutation of the Stargardt disease gene (ABCR) in age-related macular degeneration. *Science* **277**, 1805-1807 (1997).

Almasy, L. & Blangero, J. Multipoint quantitative-trait linkage analysis in general pedigrees. *Am.J.Hum.Genet.* **62**, 1198-1211 (1998).

Altmuller, J., Palmer, L. J., Fischer, G., Scherb, H., & Wjst, M. Genomewide scans of complex human diseases: True linkage is hard to find. *American Journal of Human Genetics* **69**, 936-950 (2001).

Ambati, J., Anand, A., Fernandez, S., Sakurai, E., Lynn, B. C., Kuziel, W. A., Rollins, B. J., & Ambati, B. K. An animal model of age-related macular degeneration in senescent Ccl-2- or Ccr-2-deficient mice. *Nat.Med.* **9**, 1390-1397 (2003).

Anderson, D. H., Ozaki, S., Nealon, M., Neitz, J., Mullins, R. F., Hageman, G. S., & Johnson, L. V. Local cellular sources of apolipoprotein E in the human retina and retinal pigmented epithelium: implications for the process of drusen formation. *Am J Ophthalmol.* **131**, 767-781 (2001).

Arnason, E., Sigurgislason, H., & Benedikz, E. Genetic homogeneity of Icelanders: fact or fiction? *Nature Genetics* **25**, 373-374 (2000).

Arya, R., Duggirala, R., Almasy, L., Rainwater, D. L., Mahaney, M. C., Cole, S., Dyer, T. D., Williams, K., Leach, R. J., Hixson, J. E., MacCluer, J. W., O'Connell, P., Stern, M. P., & Blangero, J. Linkage of high-density lipoprotein-cholesterol concentrations to a locus on chromosome 9p in Mexican Americans. *Nat.Genet.* **30**, 102-105 (2002).

Assini, A., Cammarata, S., Vitali, A., Colucci, M., Giliberto, L., Borghi, R., Inglese, M. L., Volpe, S., Ratto, S., Dagna-Bricarelli, F., Baldo, C., Argusti, A., Odetti, P., Piccini, A., & Tabaton, M. Plasma levels of amyloid beta-protein 42 are increased in women with mild cognitive impairment. *Neurology* **63**, 828-831 (2004).

Atwood, L. D. & Heard-Costa, N. L. Limits of fine-mapping a quantitative trait. *Genet.Epidemiol.* **24**, 99-106 (2003).

Aulchenko, Y. S., de Koning, D. J., & Haley, C. Genomewide rapid association using mixed model and regression: a fast and simple method for genomewide pedigree-based quantitative trait loci association analysis. *Genetics* **177**, 577-585 (2007).

Bailey, J. A. & Eichler, E. E. Primate segmental duplications: crucibles of evolution, diversity and disease. *Nat.Rev.Genet.* **7**, 552-564 (2006).

Baird, P. N., Guida, E., Chu, D. T., Vu, H. T. V., & Guymer, R. H. The epsilon 2 and epsilon 4 alleles of the apolipoprotein gene are associated with age-related macular degeneration. *Investigative Ophthalmology & Visual Science* **45**, 1311-1315 (2004).

Bakhru, A. & Erlinger, T. P. Smoking cessation and cardiovascular disease risk factors: results from the Third National Health and Nutrition Examination Survey. *PLoS Med.* **2**, e160 (2005).

Balding, D. J. A tutorial on statistical methods for population association studies. *Nat.Rev.Genet.* **7**, 781-791 (2006).

Baldwin, M. A. Protein identification by mass spectrometry - Issues to be considered. *Molecular & Cellular Proteomics* **3**, 1-9 (2004).

Bales, K. R., Verina, T., Cummins, D. J., Du, Y., Dodel, R. C., Saura, J., Fishman, C. E., DeLong, C. A., Piccardo, P., Petegnief, V., Ghetti, B., & Paul, S. M. Apolipoprotein E is essential for amyloid deposition in the APP(V717F) transgenic mouse model of Alzheimer's disease. *Proc.Natl.Acad.Sci.U.S A* **96**, 15233-15238 (1999).

Bamshad, M. & Wooding, S. P. Signatures of natural selection in the human genome. *Nat.Rev.Genet.* **4** , 99-111 (2003).

Barnett V. & Lewis T. (1994). Outliers in Statistical Data (3rd Edition). Wiley Publishers.

Barral, S., Francis, P. J., Schultz, D. W., Schain, M. B., Haynes, C., Majewski, J., Ott, J., Acott, T., Weleber, R. G., & Klein, M. L. Expanded genome scan in extended families with age-related macular degeneration. *Investigative Ophthalmology & Visual Science* **47**, 5453-5459 (2006).

Barrett, J. C., Fry, B., Maller, J., & Daly, M. J. Haploview: analysis and visualization of LD and haplotype maps. *Bioinformatics* **21**, 263-265 (2005).

Basun, H., Grut, M., Winblad, B., & Lannfelt, L. Apolipoprotein epsilon 4 allele and disease progression in patients with late-onset Alzheimer's disease. *Neurosci.Lett.* **183**, 32-34 (1995).

Baumann, S., Ceglarek, U., Fiedler, G. M., Lembcke, J., Leichtle, A., & Thiery, J. Standardized approach to proteome profiling of human serum based on magnetic bead separation and matrix-assisted laser desorption/ionization time-of-flight mass spectrometry. *Clin.Chem.* **51**, 973-980 (2005).

Bear, J. C., Nemec, T. F., Kennedy, J. C., Marshall, W. H., Power, A. A., Kolonel, V. M., & Burke, G. B. Inbreeding in outport Newfoundland. *Am J Med.Genet.* **29**, 649-660 (1988).

Beavis W. D. (1994). The Power and Deceit of QTL Experiments: Lessons from Comparative QTL Studies. Proceedings of the Forty-ninth Annual Corn & Sorghum Industry Conference. American Seed Trade Association. Washington, DC.

Beekman, M., Heijmans, B. T., Martin, N. G., Pedersen, N. L., Whitfield, J. B., DeFaire, U., van Baal, G. C., Snieder, H., Vogler, G. P., Slagboom, P. E., & Boomsma, D. I. Heritabilities of apolipoprotein and lipid levels in three countries. *Twin.Res* **5**, 87-97 (2002).

Bell, R. D., Sagare, A. P., Friedman, A. E., Bedi, G. S., Holtzman, D. M., Deane, R., & Zlokovic, B. V. Transport pathways for clearance of human Alzheimer's amyloid beta-peptide and apolipoproteins E and J in the mouse central nervous system. *J Cereb.Blood Flow Metab* **27**, 909-918 (2007).

Bibl, M., Esselmann, H., Otto, M., Lewczuk, P., Cepek, L., Ruther, E., Kornhuber, J., & Wiltfang, J. Cerebrospinal fluid amyloid beta peptide patterns in Alzheimer's disease patients and nondemented controls depend on sample pretreatment: Indication of carrier-mediated epitope masking of amyloid beta peptides. *Electrophoresis* **25**, 2912-2918 (2004).

Bickeboller, H., Campion, D., Brice, A., Amouyel, P., Hannequin, D., Didierjean, O., Penet, C., Martin, C., Perez-Tur, J., Michon, A., Dubois, B., Ledoze, F., Thomas-Anterion, C., Pasquier, F., Puel, M., Demonet, J. F., Moreaud, O., Babron, M. C., Meulien, D., Guez, D., Chartier-Harlin, M. C., Frebourg, T., Agid, Y., Martinez, M., & Clerget-Darpoux, F. Apolipoprotein E and Alzheimer disease: genotype-specific risks by age and sex. *Am J Hum.Genet.* **60**, 439-446 (1997).

Bihari-Varga, M., Szekely, J., & Gruber, E. Plasma high density lipoproteins in coronary, cerebral and peripheral vascular disease. The influence of various risk factors. *Atherosclerosis* **40**, 337-345 (1981).

Bird, A. C., Bressler, N. M., Bressler, S. B., Chisholm, I. H., Coscas, G., Davis, M. D., de Jong, P. T., Klaver, C. C., Klein, B. E., Klein, R., & . An international classification and grading system for age-related maculopathy and age-related macular degeneration. The International ARM Epidemiological Study Group. *Surv.Ophthalmol.* **39**, 367-374 (1995).

Bird, T. D. Genetic factors in Alzheimer's disease. *New England Journal of Medicine* **352**, 862-864 (2005).

Bjorhall, K., Miliotis, T., & Davidsson, P. Comparison of different depletion strategies for improved resolution in proteomic analysis of human serum samples. *Proteomics* **5**, 307-317 (2005).

Bjornsson, A., Gudmundsson, G., Gudfinnsson, E., Hrafnsdottir, M., Benedikz, J., Skuladottir, S., Kristjansson, K., Frigge, M. L., Kong, A., Stefansson, K., & Gulcher, J. R. Localization of a gene for migraine without aura to chromosome 4q21. *American Journal of Human Genetics* **73**, 986-993 (2003).

Blangero, J. Localization and identification of human quantitative trait loci: King Harvest has surely come. *Current Opinion in Genetics & Development* **14**, 233-240 (2004).

Blangero, J., Williams, J. T., & Almasy, L. Novel family-based approaches to genetic risk in thrombosis. *J Thromb.Haemost.* **1**, 1391-1397 (2003).

Boerwinkle, E., Chakraborty, R., & Sing, C. F. The Use of Measured Genotype Information in the Analysis of Quantitative Phenotypes in Man .1. Models and Analytical Methods. *Annals of Human Genetics* **50**, 181-194 (1986).

Bonnerjea, J., Oh, S., & Hoare, M. Protein-Purification - the Right Step at the Right Time. *Bio-Technology* **4**, 954-& (1986).

Bora, N. S., Kaliappan, S., Jha, P., Xu, Q., Sivasankar, B., Harris, C. L., Morgan, B. P., & Bora, P. S. CD59, a complement regulatory protein, controls choroidal neovascularization in a mouse model of wet-type age-related macular degeneration. *J Immunol.* **178**, 1783-1790 (2007).

Borchelt, D. R., Ratovitski, T., van Lare, J., Lee, M. K., Gonzales, V., Jenkins, N. A., Copeland, N. G., Price, D. L., & Sisodia, S. S. Accelerated amyloid deposition in the brains of transgenic mice coexpressing mutant presenilin 1 and amyloid precursor proteins. *Neuron* **19**, 939-945 (1997).

Breslow, J. L., Zannis, V. I., SanGiacomo, T. R., Third, J. L., Tracy, T., & Glueck, C. J. Studies of familial type III hyperlipoproteinemia using as a genetic marker the apoE phenotype E2/2. *J Lipid Res* **23**, 1224-1235 (1982).

Brion, J. P., Couck, A. M., Passareiro, E., & Flament-Durand, J. Neurofibrillary tangles of Alzheimer's disease: an immunohistochemical study. *J Submicrosc.Cytol.* **17**, 89-96 (1985).

Broeckel, U., Hengstenberg, C., Mayer, B., Holmer, S., Martin, L. J., Comuzzie, A. G., Blangero, J., Nurnberg, P., Reis, A., Riegger, G. A., Jacob, H. J., & Schunkert, H. A comprehensive linkage analysis for myocardial infarction and its related risk factors. *Nat.Genet.* **30**, 210-214 (2002).

Brookes, A. J. & Prince, J. A. Genetic association analysis: lessons from the study of Alzheimers Disease. *Mutation Research-Fundamental and Molecular Mechanisms of Mutagenesis* **573**, 152-159 (2005).

Callesen, A. K., Mohammed, S., Bunkenborg, J., Kruse, T. A., Cold, S., Mogensen, O., Christensen, R. D., Vach, W., Jorgensen, P. E., & Jensen, O. N. Serum protein profiling by miniaturized solid-phase extraction and matrix-assisted laser desorption/ionization mass spectrometry. *Rapid Communications in Mass Spectrometry* **19**, 1578-1586 (2005).

Camp, N. J. Genomewide transmission/disequilibrium testing - Consideration of the genotypic relative risks at disease loci. *American Journal of Human Genetics* **61**, 1424-1430 (1997).

Campbell P. N. & Smith A. D. (2000). Biochemistry Illustrated. Elsevier Health Sciences.

Campion, D., Dumanchin, C., Hannequin, D., Dubois, B., Belliard, S., Puel, M., Thomas-Anterion, C., Michon, A., Martin, C., Charbonnier, F., Raux, G., Camuzat, A., Penet, C., Mesnage, V., Martinez, M., Clerget-Darpoux, F., Brice, A., & Frebourg, T. Early-onset autosomal dominant Alzheimer disease: prevalence, genetic heterogeneity, and mutation spectrum. *Am J Hum.Genet.* **65**, 664-670 (1999).

Canning, D. R., McKeon, R. J., DeWitt, D. A., Perry, G., Wujek, J. R., Frederickson, R. C., & Silver, J. beta-Amyloid of Alzheimer's disease induces reactive gliosis that inhibits axonal outgrowth. *Exp.Neurol.* **124**, 289-298 (1993).

CANNON, G. B. The effects of natural selection on linkage disequilibrium and relative fitness in experimental populations of Drosophila melanogaster. *Genetics* **48**, 1201-1216 (1963).

Caprioli, J., Castelletti, F., Bucchioni, S., Bettinaglio, P., Bresin, E., Pianetti, G., Gamba, S., Brioschi, S., Daina, E., Remuzzi, G., & Noris, M. Complement factor H mutations and gene polymorphisms in haemolytic uraemic syndrome: the C-257T, the A2089G and the G2881T polymorphisms are strongly associated with the disease. *Hum.Mol.Genet.* **12**, 3385-3395 (2003).

Carlson, C. S., Aldred, S. F., Lee, P. K., Tracy, R. P., Schwartz, S. M., Rieder, M., Liu, K. A., Williams, O. D., Iribarren, C., Lewis, E. C., Fornage, M., Boerwinkle, E., Gross, M., Jaquish, C., Nickerson, D. A., Myers, R. M., Siscovick, D. S., & Reiner, A. P. Polymorphisms within the C-reactive protein (CRP) promoter region are associated with plasma CRP levels. *American Journal of Human Genetics* **77**, 64-77 (2005).

Caulfield, M., Munroe, P., Pembroke, J., Samani, N., Dominiczak, A., Brown, M., Benjamin, N., Webster, J., Ratcliffe, P., O'Shea, S., Papp, J., Taylor, E., Dobson, R., Knight, J., Newhouse, S., Hooper, J., Lee, W., Brain, N., Clayton, D., Lathrop, G. M., Farrall, M., & Connell, J. Genome-wide mapping of human loci for essential hypertension. *Lancet* **361**, 2118-2123 (2003).

Chait, B. T. Mass spectrometry: Bottom-up or top-down? *Science* **314**, 65-66 (2006).

Chalmers, M. J., Mackay, C. L., Hendrickson, C. L., Wittke, S., Walden, M., Mischak, H., Fliser, D., Just, I., & Marshall, A. G. Combined top-down and bottom-up mass spectrometric approach to characterization of biomarkers for renal disease. *Anal.Chem.* **77**, 7163-7171 (2005).

Chaney, M. O., Webster, S. D., Kuo, Y. M., & Roher, A. E. Molecular modeling of the A beta 1-42 peptide from Alzheimer's disease. *Protein Engineering* **11**, 761-767 (1998).

Charlesworth B. & Hughes K. A. (2000). The maintenance of Genetic Variation in Life-history Traits. Cambridge University Press, Cambridge.

Chelius, D. & Bondarenko, P. V. Quantitative profiling of proteins in complex mixtures using liquid chromatography and mass spectrometry. *Journal of Proteome Research* **1**, 317-323 (2002).

Chen, L. J., Liu, D. T., Tam, P. O., Chan, W. M., Liu, K., Chong, K. K., Lam, D. S., & Pang, C. P. Association of complement factor H polymorphisms with exudative age-related macular degeneration. *Mol.Vis.* **12**, 1536-1542 (2006).

Chernokalskaya, E., Gutierrez, S., Pitt, A. M., & Leonard, J. T. Ultrafiltration for proteomic sample preparation. *Electrophoresis* **25**, 2461-2468 (2004).

Chinnery P. F. (2006). Neuroscience for neurologists. Imperial College Press.

Christen, W. G., Glynn, R. J., Manson, J. E., Ajani, U. A., & Buring, J. E. A prospective study of cigarette smoking and risk of age-related macular degeneration in men. *JAMA* **276** , 1147-1151 (1996).

Clark, A. G. The role of haplotypes in candidate gene studies. *Genet.Epidemiol.* **27**, 321-333 (2004).

Clark, S. J., Higman, V. A., Mulloy, B., Perkins, S. J., Lea, S. M., Sim, R. B., & Day, A. J. His-384 allotypic variant of factor H associated with age-related macular degeneration has different heparin binding properties from the non-disease-associated form. *J Biol.Chem.* **281**, 24713-24720 (2006).

Coffey, P. J., Gias, C., McDermott, C. J., Lundh, P., Pickering, M. C., Sethi, C., Bird, A., Fitzke, F. W., Maass, A., Chen, L. L., Holder, G. E., Luthert, P. J., Salt, T. E., Moss, S. E., & Greenwood, J. Complement factor H deficiency in aged mice causes retinal abnormalities and visual dysfunction. *Proc.Natl.Acad.Sci.U.S A* **104**, 16651-16656 (2007).

Conley, Y. P., Jakobsdottir, J., Mah, T., Weeks, D. E., Klein, R., Kuller, L., Ferrell, R. E., & Gorin, M. B. CFH, ELOVL4, PLEKHA1 and LOC387715 genes and susceptibility to age-related maculopathy: AREDS and CHS cohorts and meta-analyses. *Hum.Mol.Genet.* **15**, 3206-3218 (2006).

Conrad, D. F., Andrews, T. D., Carter, N. P., Hurles, M. E., & Pritchard, J. K. A high-resolution survey of deletion polymorphism in the human genome. *Nat.Genet.* **38**, 75-81 (2006).

Corbo, R. M., Scacchi, R., Mureddu, L., Mulas, G., & Alfano, G. Apolipoprotein E polymorphism in Italy investigated in native plasma by a simple polyacrylamide gel isoelectric focusing technique. Comparison with frequency data of other European populations. *Ann.Hum.Genet.* **59**, 197-209 (1995).

Corbo, R. M., Scacchi, R., Mureddu, L., Mulas, G., Castrechini, S., & Rivasi, P. Apolipoprotein B, apolipoprotein E, and angiotensin-converting enzyme polymorphisms in 2 Italian populations at different risk for coronary artery disease and comparison of allele frequencies among European populations. *Human Biology* **71**, 933-945 (1999).

Corder, E. H., Saunders, A. M., Risch, N. J., Strittmatter, W. J., Schmechel, D. E., Gaskell, P. C., Rimmler, J. B., Locke, P. A., Conneally, P. M., Schmader, K. E., Small, G. W.,

Roses, A. D., Haines, J. L., & Pericakvance, M. A. Protective Effect of Apolipoprotein-e Type-2 Allele for Late-Onset Alzheimer-Disease. *Nature Genetics* **7**, 180-184 (1994).

Corder, E. H., Saunders, A. M., Strittmatter, W. J., Schmechel, D. E., Gaskell, P. C., Small, G. W., Roses, A. D., Haines, J. L., & Pericakvance, M. A. Gene Dose of Apolipoprotein-e Type-4 Allele and the Risk of Alzheimers-Disease in Late-Onset Families. *Science* **261**, 921-923 (1993).

Crabb, J. W., Miyagi, M., Gu, X. R., Shadrach, K., West, K. A., Sakaguchi, H., Kamei, M., Hasan, A., Yan, L., Rayborn, M. E., Salomon, R. G., & Hollyfield, J. G. Drusen proteome analysis: An approach to the etiology of age-related macular degeneration. *Proceedings of the National Academy of Sciences of the United States of America* **99**, 14682-14687 (2002).

Crow J. F. & Kimura M. (1970). An Introduction to Population Genetics Theory. Harper and Row, New York.

Davey, S. G. & Ebrahim, S. 'Mendelian randomization': can genetic epidemiology contribute to understanding environmental determinants of disease? *Int.J Epidemiol.* **32**, 1-22 (2003).

Davey, S. G., Ebrahim, S., Lewis, S., Hansell, A. L., Palmer, L. J., & Burton, P. R. Genetic epidemiology and public health: hope, hype, and future prospects. *Lancet* **366**, 1484-1498 (2005).

Davignon, J., Gregg, R. E., & Sing, C. F. Apolipoprotein-e Polymorphism and Atherosclerosis. *Arteriosclerosis* **8**, 1-21 (1988).

Daw, E. W., Payami, H., Nemens, E. J., Nochlin, D., Bird, T. D., Schellenberg, G. D., & Wijsman, E. M. The number of trait loci in late-onset Alzheimer disease. *American Journal of Human Genetics* **66**, 196-204 (2000).

Dawn, T. M. & Barrett, J. H. Genetic linkage studies. *Lancet* **366**, 1036-1044 (2005).

de Cordoba, S. R., Esparza-Gordillo, J., de Jorge, E. G., Lopez-Trascasa, M., & Sanchez-Corral, P. The human complement factor H: functional roles, genetic variations and disease associations. *Molecular Immunology* **41**, 355-367 (2004).

de Noo, M. E., Tollenaar, R. A., Ozalp, A., Kuppen, P. J., Bladergroen, M. R., Eilers, P. H., & Deelder, A. M. Reliability of human serum protein profiles generated with C8 magnetic beads assisted MALDI-TOF mass spectrometry. *Anal.Chem.* **77**, 7232-7241 (2005).

DeCarlo, L. T. On the meaning and use of kurtosis. *Psychological Methods* **2**, 292-307 (1997).

DeMattos, R. B., Bales, K. R., Cummins, D. J., Paul, S. M., & Holtzman, D. M. Brain to plasma amyloid-beta efflux: a measure of brain amyloid burden in a mouse model of Alzheimer's disease. *Science* **295**, 2264-2267 (2002).

DeSouza, L., Diehl, G., Rodrigues, M. J., Guo, J. Z., Romaschin, A. D., Colgan, T. J., & Siu, K. W. M. Search for cancer markers from endometrial tissues using differentially labeled

tags iTRAQ and clCAT with multidimensional liquid chromatography and tandem mass spectrometry. *Journal of Proteome Research* **4**, 377-386 (2005).

Despriet, D. D. G., Klaver, C. C. W., Witteman, J. C. M., Bergen, A. A. B., Kardys, I., de Maat, M. P. M., Boekhoorn, S. S., Vingerling, J. R., Hofman, A., Oostra, B. A., Uitterlinden, A. G., Stijnen, T., van Duijn, C. M., & De Jong, P. T. V. M. Complement factor H polymorphism, complement activators, and risk of age-related macular degeneration. *Jama-Journal of the American Medical Association* **296**, 301-309 (2006).

Devlin, B. & Roeder, K. Genomic control for association studies. *Biometrics* **55**, 997-1004 (1999).

Dewan, A., Liu, M. G., Hartman, S., Zhang, S. S. M., Liu, D. T. L., Zhao, C., Tam, P. O. S., Chan, W. M., Lam, D. S. C., Snyder, M., Barnstable, C., Pang, C. P., & Hoh, J. HTRA1 promoter polymorphism in wet age-related macular degeneration. *Science* **314**, 989-992 (2006).

Di Rienzo, A. Population genetics models of common diseases. *Current Opinion in Genetics & Development* **16**, 630-636 (2006).

Dickerson, T. J. & Janda, K. D. Glycation of the amyloid beta-protein by a nicotine metabolite: a fortuitous chemical dynamic between smoking and Alzheimer's disease. *Proc.Natl.Acad.Sci.U.S A* **100**, 8182-8187 (2003).

DiScipio, R. G., Daffern, P. J., Schraufstatter, I. U., & Sriramarao, P. Human polymorphonuclear leukocytes adhere to complement factor H through an interaction that involves alphaMbeta2 (CD11b/CD18). *J Immunol.* **160**, 4057-4066 (1998).

Dodart, J. C., Marr, R. A., Koistinaho, M., Gregersen, B. M., Malkani, S., Verma, I. M., & Paul, S. M. Gene delivery of human apolipoprotein E alters brain A beta burden in a mouse model of Alzheimer's disease. *Proceedings of the National Academy of Sciences of the United States of America* **102**, 1211-1216 (2005).

Dong, L. M., Wilson, C., Wardell, M. R., Simmons, T., Mahley, R. W., Weisgraber, K. H., & Agard, D. A. Human apolipoprotein E. Role of arginine 61 in mediating the lipoprotein preferences of the E3 and E4 isoforms. *J Biol.Chem.* **269**, 22358-22365 (1994).

Dudal, S., Krzywkowski, P., Paquette, J., Morissette, C., Lacombe, D., Tremblay, P., & Gervais, F. Inflammation occurs early during the Abeta deposition process in TgCRND8 mice. *Neurobiol.Aging* **25**, 861-871 (2004).

Duering, M., Grimm, M. O. W., Grimm, H. S., Schroder, J., & Hartmann, T. Mean age of onset in familial Alzheimer's disease is determined by amyloid beta 42. *Neurobiology of Aging* **26**, 785-788 (2005).

Edmondson, J. C., Liem, R. K., Kuster, J. E., & Hatten, M. E. Astrotactin: a novel neuronal cell surface antigen that mediates neuron-astroglial interactions in cerebellar microcultures. *J Cell Biol.* **106**, 505-517 (1988).

Edwards, A. O. & Malek, G. Molecular genetics of AMD and current animal models. *Angiogenesis.* **10**, 119-132 (2007).

Edwards, A. O., Ritter, R., Abel, K. J., Manning, A., Panhuysen, C., & Farrer, L. A. Complement factor H polymorphism and age-related macular degeneration. *Science* **308**, 421-424 (2005).

Edwards, K. L., Mahaney, M. C., Motulsky, A. G., & Austin, M. A. Pleiotropic genetic effects on LDL size, plasma triglyceride, and HDL cholesterol in families. *Arterioscler.Thromb.Vasc.Biol.* **19**, 2456-2464 (1999).

Eichner, J. E., Dunn, S. T., Perveen, G., Thompson, D. M., Stewart, K. E., & Stroehla, B. C. Apolipoprotein E polymorphism and cardiovascular disease: A HuGE review. *American Journal of Epidemiology* **155**, 487-495 (2002).

Ellis, W. S. & Starmer, W. T. Inbreeding as measured by isonymy, pedigrees, and population size in Torbel, Switzerland. *Am J Hum.Genet.* **30**, 366-376 (1978).

Engvall, E. & Perlmann, P. Enzyme-Linked Immunosorbent Assay (Elisa) Quantitative Assay of Immunoglobulin-G. *Immunochemistry* **8**, 871-& (1971).

Ertekin-Taner, N., Allen, M., Fadale, D., Scanlin, L., Younkin, L., Petersen, R. C., Graff-Radford, N., & Younkin, S. G. Genetic variants in a haplotype block spanning IDE are significantly associated with plasma Abeta42 levels and risk for Alzheimer disease. *Hum.Mutat.* **23**, 334-342 (2004).

Ertekin-Taner, N., Graff-Radford, N., Younkin, L. H., Eckman, C., Baker, M., Adamson, J., Ronald, J., Blangero, J., Hutton, M., & Younkin, S. G. Linkage of plasma A beta 42 to a quantitative locus on chromosome 10 in late-onset Alzheimer's disease pedigrees. *Science* **290**, 2303-+ (2000).

Ertekin-Taner, N., Graff-Radford, N., Younkin, L. H., Eckman, C., Adamson, J., Schaid, D. J., Blangero, J., Hutton, M., & Younkin, S. G. Heritability of plasma amyloid beta in typical late-onset Alzheimer's disease pedigrees. *Genetic Epidemiology* **21**, 19-30 (2001).

Ertekin-Taner, N., Ronald, J., Asahara, H., Younkin, L., Hella, M., Jain, S., Gnida, E., Younkin, S., Fadale, D., Ohyagi, Y., Singleton, A., Scanlin, L., de Andrade, M., Petersen, R., Graff-Radford, N., Hutton, M., & Younkin, S. Fine mapping of the alpha-T catenin gene to a quantitative trait locus on chromosome 10 in late-onset Alzheimer's disease pedigrees. *Human Molecular Genetics* **12**, 3133-3143 (2003).

Ertekin-Taner, N., Ronald, J., Feuk, L., Prince, J., Tucker, M., Younkin, L., Hella, M., Jain, S., Hackett, A., Scanlin, L., Kelly, J., Kihiko-Ehman, M., Neltner, M., Hersh, L., Kindy, M., Markesbery, W., Hutton, M., de Andrade, M., Petersen, R. C., Graff-Radford, N., Estus, S., Brookes, A. J., & Younkin, S. G. Elevated amyloid beta protein (A beta 42) and late onset Alzheimer's disease are associated with single nucleotide polymorphisms in the urokinase-type plasminogen activator gene. *Human Molecular Genetics* **14**, 447-460 (2005).

Ertekin-Taner, N., Younkin, L. H., Yager, D. M., Parfitt, F., Baker, M. C., Asthana, S., Hutton, M. L., Younkin, S. G., & Graff-Radford, N. R. Plasma amyloid beta protein is elevated in late-onset Alzheimer disease families. *Neurology* **70**, 596-606 (2008).

Escamilla, M. A. Population isolates: their special value for locating genes for bipolar disorder. *Bipolar Disorders* **3**, 299-317 (2001).

Esparza-Gordillo, J., Soria, J. M., Buil, A., Almasy, L., Blangero, J., Fontcuberta, J., & de Cordoba, S. R. Genetic and environmental factors influencing the human factor H plasma levels. *Immunogenetics* **56**, 77-82 (2004).

Eudy, J. D., Weston, M. D., Yao, S., Hoover, D. M., Rehm, H. L., Ma-Edmonds, M., Yan, D., Ahmad, I., Cheng, J. J., Ayuso, C., Cremers, C., Davenport, S., Moller, C., Talmadge, C. B., Beisel, K. W., Tamayo, M., Morton, C. C., Swaroop, A., Kimberling, W. J., & Sumegi, J. Mutation of a gene encoding a protein with extracellular matrix motifs in Usher syndrome type IIa. *Science* **280**, 1753-1757 (1998).

Evans, J. R. Risk factors for age-related macular degeneration. *Prog.Retin.Eye Res* **20**, 227-253 (2001).

Falchi, M., Forabosco, P., Mocci, E., Borlino, C. C., Picciau, A., Virdis, E., Persico, I., Parracciani, D., Angius, A., & Pirastu, M. A genomewide search using an original pairwise sampling approach for large genealogies identifies a new locus for total and low-density lipoprotein cholesterol in two genetically differentiated isolates of Sardinia. *American Journal of Human Genetics* **75**, 1015-1031 (2004).

Fauser, S., Luberichs, J., & Schuttauf, F. Genetic animal models for retinal degeneration. *Surv.Ophthalmol.* **47**, 357-367 (2002).

Fearon, D. T. Regulation by membrane sialic acid of beta1H-dependent decay-dissociation of amplification C3 convertase of the alternative complement pathway. *Proc.Natl.Acad.Sci.U.S A* **75**, 1971-1975 (1978).

Fenn, J. B., Mann, M., Meng, C. K., Wong, S. F., & Whitehouse, C. M. Electrospray Ionization for Mass-Spectrometry of Large Biomolecules. *Science* **246**, 64-71 (1989).

Field A. (2000). Discovering Statistics, using SPSS for Windows. SAGE Publications.

Finckh, U. The future of genetic association studies in Alzheimer disease. *Journal of Neural Transmission* **110**, 253-266 (2003).

Finckh, U., Kuschel, C., Anagnostouli, M., Patsouris, E., Pantes, G. V., Gatzonis, S., Kapaki, E., Davaki, P., Lamszus, K., Stavrou, D., & Gal, A. Novel mutations and repeated findings of mutations in familial Alzheimer disease. *Neurogenetics* **6**, 85-89 (2005).

Fisher R. A. (1958) The Genetical Theory Natural Selection (2nd Edition). Dover Publications, New York.

Fisher, S. A., Abecasis, G. R., Yashar, B. M., Zareparsi, S., Swaroop, A., Iyengar, S. K., Klein, B. E. K., Klein, R., Lee, K. E., Majewski, J., Schultz, D. W., Klein, M. L., Seddon, J. M., Santangelo, S. L., Weeks, D. E., Conley, Y. P., Mah, T. S., Schmidt, S., Haines, J. L., Pericak-Vance, M. A., Gorin, M. B., Schulz, H. L., Pardi, F., Lewis, C. M., & Weber, B. H. F. Meta-analysis of genome scans of age-related macular degeneration. *Human Molecular Genetics* **14**, 2257-2264 (2005).

Forenbaher, S. The earliest islanders of the eastern Adriatic. *Coll.Antropol.* **23**, 521-530 (1999).

Francis, P. J., Schultz, D. W., Hamon, S., Ott, J., Weleber, R. G., & Klein, M. L. Haplotypes in the Complement Factor H (CFH) Gene: Associations with Drusen and Advanced Age-Related Macular Degeneration. *PLoS ONE.* **2**, e1197 (2007).

Freeman, D. J., Griffin, B. A., Murray, E., Lindsay, G. M., Gaffney, D., Packard, C. J., & Shepherd, J. Smoking and plasma lipoproteins in man: effects on low density lipoprotein cholesterol levels and high density lipoprotein subfraction distribution. *Eur.J Clin.Invest* **23**, 630-640 (1993).

Friedman, D. S., O'Colmain, B. J., Munoz, B., Tomany, S. C., McCarty, C., de Jong, P. T., Nemesure, B., Mitchell, P., & Kempen, J. Prevalence of age-related macular degeneration in the United States. *Arch.Ophthalmol.* **122**, 564-572 (2004).

Frohlich, T. & Arnold, G. J. Proteome research based on modern liquid chromatography - tandem mass spectrometry: separation, identification and quantification. *Journal of Neural Transmission* **113**, 973-994 (2006).

Fu, L., Garland, D., Yang, Z., Shukla, D., Rajendran, A., Pearson, E., Stone, E. M., Zhang, K., & Pierce, E. A. The R345W mutation in EFEMP1 is pathogenic and causes AMD-like deposits in mice. *Hum.Mol.Genet.* **16**, 2411-2422 (2007).

Fujii, K., Nakano, T., Kanazawa, M., Akimoto, S., Hirano, T., Kato, H., & Nishimura, T. Clinical-scale high-throughput human plasma proteome clinical analysis: Lung adenocarcinoma. *Proteomics* **5**, 1150-1159 (2005).

Fukumoto, H., Tennis, M., Locascio, J. J., Hyman, B. T., Growdon, J. H., & Irizarry, M. C. Age but not diagnosis is the main predictor of plasma amyloid beta-protein levels. *Archives of Neurology* **60**, 958-964 (2003).

Fuse, N., Miyazawa, A., Mengkegale, M., Yoshida, M., Wakusawa, R., Abe, T., & Tamai, M. Polymorphisms in Complement Factor H and Hemicentin-1 genes in a Japanese population with dry-type age-related macular degeneration. *Am J Ophthalmol.* **142**, 1074-1076 (2006).

Games, D., Adams, D., Alessandrini, R., Barbour, R., Berthelette, P., Blackwell, C., Carr, T., Clemens, J., Donaldson, T., Gillespie, F., & . Alzheimer-type neuropathology in transgenic mice overexpressing V717F beta-amyloid precursor protein. *Nature* **373**, 523-527 (1995).

Gatz, M., Pedersen, N. L., Berg, S., Johansson, B., Johansson, K., Mortimer, J. A., Posner, S. F., Viitanen, M., Winblad, B., & Ahlbom, A. Heritability for Alzheimer's disease: The study of dementia in Swedish twins. *Journals of Gerontology Series A-Biological Sciences and Medical Sciences* **52**, M117-M125 (1997).

Gay-Bellile, C., Bengoufa, D., Houze, P., Le Carrer, D., Benlakehal, M., Bousquet, B., Gourmel, B., & Le Bricon, T. Automated multicapillary electrophoresis for analysis of human serum proteins. *Clinical Chemistry* **49**, 1909-1915 (2003).

Genin, E. & ClergetDarpoux, F. Association studies in consanguineous populations. *American Journal of Human Genetics* **58**, 861-866 (1996).

Geno, P. W. & Macfarlane, R. D. Secondary-Electron Emission Induced by Impact of Low-Velocity Molecular-Ions on A Microchannel Plate. *International Journal of Mass Spectrometry and Ion Processes* **92**, 195-210 (1989).

Gianfrancesco, F., Esposito, T., Ombra, M. N., Forabosco, P., Maninchedda, G., Fattorini, M., Casula, S., Vaccargiu, S., Casu, G., Cardia, F., Deiana, I., Melis, P., Falchi, M., & Pirastu, M. Identification of a novel gene and a common variant associated with uric acid nephrolithiasis in a sardinian genetic isolate. *American Journal of Human Genetics* **72**, 1479-1491 (2003).

Glenner, G. G. & Wong, C. W. Alzheimers-Disease and Downs-Syndrome - Sharing of A Unique Cerebrovascular Amyloid Fibril Protein. *Biochemical and Biophysical Research Communications* **122**, 1131-1135 (1984).

Gnjec, A., Fonte, J. A., Atwood, C., & Martins, R. N. Transition metal chelator therapy--a potential treatment for Alzheimer's disease? *Front Biosci.* **7**, d1016-d1023 (2002).

Gold, B., Merriam, J. E., Zernant, J., Hancox, L. S., Taiber, A. J., Gehrs, K., Cramer, K., Neel, J., Bergeron, J., Barile, G. R., Smith, R. T., Dean, M., & Allikmets, R. Variation in factor B (BF) and complement component 2 (C2) genes is associated with age-related macular degeneration. *Nature Genetics* **38**, 458-462 (2006).

Goldstein, D. B. Islands of linkage disequilibrium. *Nat.Genet.* **29**, 109-111 (2001).

Gotoh, N., Yamada, R., Hiratani, H., Renault, V., Kuroiwa, S., Monet, M., Toyoda, S., Chida, S., Mandai, M., Otani, A., Yoshimura, N., & Matsuda, F. No association between complement factor H gene polymorphism and exudative age-related macular degeneration in Japanese. *Hum.Genet.* **120**, 139-143 (2006).

Gotz, J., Probst, A., Spillantini, M. G., Schafer, T., Jakes, R., Burki, K., & Goedert, M. Somatodendritic localization and hyperphosphorylation of tau protein in transgenic mice expressing the longest human brain tau isoform. *EMBO J* **14**, 1304-1313 (1995).

Graff-Radford, N. R., Crook, J. E., Lucas, J., Boeve, B. F., Knopman, D. S., Ivnik, R. J., Smith, G. E., Younkin, L. H., Petersen, R. C., & Younkin, S. G. Association of low plasma Abeta42/Abeta40 ratios with increased imminent risk for mild cognitive impairment and Alzheimer disease. *Arch.Neurol.* **64**, 354-362 (2007).

Granger, J., Siddiqui, J., Copeland, S., & Remick, D. Albumin depletion of human plasma also removes low abundance proteins including the cytokines. *Proteomics* **5**, 4713-4718 (2005).

Gravina, S. A., Ho, L. B., Eckman, C. B., Long, K. E., Otvos, L., Younkin, L. H., Suzuki, N., & Younkin, S. G. Amyloid-Beta Protein (A-Beta) in Alzheimers-Disease Brain - Biochemical and Immunocytochemical Analysis with Antibodies Specific for Forms Ending at A-Beta-40 Or A-Beta-42(43). *Journal of Biological Chemistry* **270**, 7013-7016 (1995).

Gretarsdottir, S., Sveinbjornsdottir, S., Jonsson, H. H., Jakobsson, F., Einarsdottir, E., Agnarsson, U., Shkolny, D., Einarsson, G., Gudjonsdottir, H. M., Valdimarsson, E. M., Einarsson, O. B., Thorgeirsson, G., Hadzic, R., Jonsdottir, S., Reynisdottir, S. T., Bjarnadottir, S. M., Gudmundsdottir, T., Gudlaugsdottir, G. J., Gill, R., Lindpaintner, K., Sainz, J., Hannesson, H. H., Sigurdsson, G. T., Frigge, M. L., Kong, A., Gudnason, V.,

Stefansson, K., & Gulcher, J. R. Localization of a susceptibility gene for common forms of stroke to 5q12. *American Journal of Human Genetics* **70**, 593-603 (2002).

Gretarsdottir, S., Thorleifsson, G., Reynisdottir, S. T., Manolescu, A., Jonsdottir, S., Jonsdottir, T., Gudmundsdottir, T., Bjarnadottir, S. M., Einarsson, O. B., Gudjonsdottir, H. M., Hawkins, M., Gudmundsson, G., Gudmundsdottir, H., Andrason, H., Gudmundsdottir, A. S., Sigurdardottir, M., Chou, T. T., Nahmias, J., Goss, S., Sveinbjornsdottir, S., Valdimarsson, E. M., Jakobsson, F., Agnarsson, U., Gudnason, V., Thorgeirsson, G., Fingerle, J., Gurney, M., Gudbjartsson, D., Frigge, M. L., Kong, A., Stefansson, K., & Gulcher, J. R. The gene encoding phosphodiesterase 4D confers risk of ischemic stroke. *Nature Genetics* **35**, 131-138 (2003).

Gudbjartsson, D. F., Arnar, D. O., Helgadottir, A., Gretarsdottir, S., Holm, H., Sigurdsson, A., Jonasdottir, A., Baker, A., Thorleifsson, G., Kristjansson, K., Palsson, A., Blondal, T., Sulem, P., Backman, V. M., Hardarson, G. A., Palsdottir, E., Helgason, A., Sigurjonsdottir, R., Sverrisson, J. T., Kostulas, K., Ng, M. C. Y., Baum, L., So, W. Y., Wong, K. S., Chan, J. C. N., Furie, K. L., Greenberg, S. M., Sale, M., Kelly, P., Macrae, C. A., Smith, E. E., Rosand, J., Hillert, J., Ma, R. C. W., Ellinor, P. T., Thorgeirsson, G., Gulcher, J. R., Kong, A., Thorsteinsdottir, U., & Stefansson, K. Variants conferring risk of atrial fibrillation on chromosome 4q25. *Nature* **448**, 353-3U9 (2007).

Gudmundsson, J., Sulem, P., Manolescu, A., Amundadottir, L. T., Gudbjartsson, D., Helgason, A., Rafnar, T., Bergthorsson, J. T., Agnarsson, B. A., Baker, A., Sigurdsson, A., Benediktsdottir, K. R., Jakobsdottir, M., Xu, J. F., Blondal, T., Kostic, J., Sun, J. L., Ghosh, S., Stacey, S. N., Mouy, M., Saemundsdottir, J., Backman, V. M., Kristjansson, K., Tres, A., Partin, A. W., Albers-Akkers, M. T., Marcos, J. G. I., Walsh, P. C., Swinkels, D. W., Navarrete, S., Isaacs, S. D., Aben, K. K., Graif, T., Cashy, J., Ruiz-Echarri, M., Wiley, K. E., Suarez, B. K., Witjes, J. A., Frigge, M., Ober, C., Jonsson, E., Einarsson, G. V., Mayordomo, J. I., Kiemeney, L. A., Isaacs, W. B., Catalona, W. J., Barkardottir, R. B., Gulcher, J. R., Thorsteinsdottir, U., Kong, A., & Stefansson, K. Genome-wide association study identifies a second prostate cancer susceptibility variant at 8q24. *Nature Genetics* **39**, 631-637 (2007).

Gudmundsson, J., Sulem, P., Rafnar, T., Bergthorsson, J. T., Manolescu, A., Gudbjartsson, D., Agnarsson, B. A., Sigurdsson, A., Benediktsdottir, K. R., Blondal, T., Jakobsdottir, M., Stacey, S. N., Kostic, J., Kristinsson, K. T., Birgisdottir, B., Ghosh, S., Magnusdottir, D. N., Thorlacius, S., Thorleifsson, G., Zheng, S. L., Sun, J., Chang, B. L., Elmore, J. B., Breyer, J. P., McReynolds, K. M., Bradley, K. M., Yaspan, B. L., Wiklund, F., Stattin, P., Lindstrom, S., Adami, H. O., McDonnell, S. K., Schaid, D. J., Cunningham, J. M., Wang, L., Cerhan, J. R., St Sauver, J. L., Isaacs, S. D., Wiley, K. E., Partin, A. W., Walsh, P. C., Polo, S., Ruiz-Echarri, M., Navarrete, S., Fuertes, F., Saez, B., Godino, J., Weijerman, P. C., Swinkels, D. W., Aben, K. K., Witjes, J. A., Suarez, B. K., Helfand, B. T., Frigge, M. L., Kristjansson, K., Ober, C., Jonsson, E., Einarsson, G. V., Xu, J. F., Gronberg, H., Smith, J. R., Thibodeau, S. N., Isaacs, W. B., Catalona, W. J., Mayordomo, J. I., Kiemeney, L. A., Barkardottir, R. B., Gulcher, J. R., Thorsteinsdottir, U., Kong, A., & Stefansson, K. Common sequence variants on 2p15 and Xp11.22 confer susceptibility to prostate cancer. *Nature Genetics* **40**, 281-283 (2008).

Gulcher, J. R., Jonsson, P., Kong, A., Kristjansson, K., Frigge, M. L., Karason, A., Einarsdottir, I. E., Stefansson, H., Einarsdottir, A. S., Sigurdardottir, S., Baldursson, S., Bjornsdottir, S., Hrafnkelsdottir, S. M., Jakobsson, F., Benedickz, J., & Stefansson, K.

Mapping of a familial essential tremor gene, FET1, to chromosome 3q13. *Nature Genetics* **17**, 84-87 (1997).

Guo, S. W. & Thompson, E. A. Performing the exact test of Hardy-Weinberg proportion for multiple alleles. *Biometrics* **48**, 361-372 (1992).

Gygi, S. P., Rist, B., Gerber, S. A., Turecek, F., Gelb, M. H., & Aebersold, R. Quantitative analysis of complex protein mixtures using isotope-coded affinity tags. *Nature Biotechnology* **17**, 994-999 (1999).

Hageman, G. S., Anderson, D. H., Johnson, L. V., Hancox, L. S., Taiber, A. J., Hardisty, L. I., Hageman, J. L., Stockman, H. A., Borchardt, J. D., Gehrs, K. M., Smith, R. J., Silvestri, G., Russell, S. R., Klaver, C. C., Barbazetto, I., Chang, S., Yannuzzi, L. A., Barile, G. R., Merriam, J. C., Smith, R. T., Olsh, A. K., Bergeron, J., Zernant, J., Merriam, J. E., Gold, B., Dean, M., & Allikmets, R. A common haplotype in the complement regulatory gene factor H (HF1/CFH) predisposes individuals to age-related macular degeneration. *Proc.Natl.Acad.Sci.U.S A* **102**, 7227-7232 (2005).

Hageman, G. S., Hancox, L. S., Taiber, A. J., Gehrs, K. M., Anderson, D. H., Johnson, L. V., Radeke, M. J., Kavanagh, D., Richards, A., Atkinson, J., Meri, S., Bergeron, J., Zernant, J., Merriam, J., Gold, B., Allikmets, R., & Dean, M. Extended haplotypes in the complement factor H (CFH) and CFH-related (CFHR) family of genes protect against age-related macular degeneration: characterization, ethnic distribution and evolutionary implications. *Ann.Med.* **38**, 592-604 (2006).

Hageman, G. S., Luthert, P. J., Victor Chong, N. H., Johnson, L. V., Anderson, D. H., & Mullins, R. F. An integrated hypothesis that considers drusen as biomarkers of immune-mediated processes at the RPE-Bruch's membrane interface in aging and age-related macular degeneration. *Prog.Retin.Eye Res* **20**, 705-732 (2001).

Hageman, G. S., Mullins, R. F., Russell, S. R., Johnson, L. V., & Anderson, D. H. Vitronectin is a constituent of ocular drusen and the vitronectin gene is expressed in human retinal pigmented epithelial cells. *FASEB J* **13**, 477-484 (1999).

Haines, J. L., Hauser, M. A., Schmidt, S., Scott, W. K., Olson, L. M., Gallins, P., Spencer, K. L., Kwan, S. Y., Noureddine, M., Gilbert, J. R., Schnetz-Boutaud, N., Agarwal, A., Postel, E. A., & Pericak-Vance, M. A. Complement factor H variant increases the risk of age-related macular degeneration. *Science* **308**, 419-421 (2005).

Haines, J. L., Schnetz-Boutaud, N., Schmidt, S., Scott, W. K., Agarwal, A., Postel, E. A., Olson, L., Kenealy, S. J., Hauser, M., Gilbert, J. R., & Pericak-Vance, M. A. Functional candidate genes in age-related macular degeneration: significant association with VEGF, VLDLR, and LRP6. *Invest Ophthalmol.Vis.Sci.* **47**, 329-335 (2006).

Hammond, C. J., Webster, A. R., Snieder, H., Bird, A. C., Gilbert, C. E., & Spector, T. D. Genetic influence on early age-related maculopathy: a twin study. *Ophthalmology* **109**, 730-736 (2002).

Hanlon, C. S. & Rubinsztein, D. C. Arginine Residues at Codon-112 and Codon-158 in the Apolipoprotein-e Gene Correspond to the Ancestral State in Humans. *Atherosclerosis* **112**, 85-90 (1995).

Hardy, J. & Selkoe, D. J. Medicine - The amyloid hypothesis of Alzheimer's disease: Progress and problems on the road to therapeutics. *Science* **297**, 353-356 (2002).

Harris, R. D., Nindl, G., Balcavage, W. X., Weiner, W., & Johnson, M. T. Use of proteomics methodology to evaluate inflammatory protein expression in tendinitis. *Biomed.Sci.Instrum.* **39**, 493-499 (2003).

Hartl D. L. & Clark A. G. (1997). Principles of Population Genetics (3rd Edition). Sinauer Associates, Inc. Sunderland, MA.

Hatten, M. E. & Mason, C. A. Mechanisms of glial-guided neuronal migration in vitro and in vivo. *Experientia* **46**, 907-916 (1990).

Hayes, R. N. & Gross, M. L. Collision-Induced Dissociation. *Methods in Enzymology* **193**, 237-263 (1990).

Heath, S. C. Markov chain Monte Carlo segregation and linkage analysis for oligogenic models. *Am J Hum.Genet.* **61**, 748-760 (1997).

Helgadottir, A., Manolescu, A., Thorleifsson, G., Gretarsdottir, S., Jonsdottir, H., Thorsteinsdottir, U., Samani, N. J., Gudmundsson, G., Grant, S. F. A., Thorgeirsson, G., Sveinbjornsdottir, S., Valdimarsson, E. M., Matthiasson, S. E., Johannsson, H., Gudmundsdottir, O., Gurney, M. E., Sainz, J., Thorhallsdottir, M., Andresdottir, M., Frigge, M. L., Topol, E. J., Kong, A., Gudnason, V., Hakonarson, H., Gulcher, J. R., & Stefansson, K. The gene encoding 5-lipoxygenase activating protein confers risk of myocardial infarction and stroke. *Nature Genetics* **36**, 233-239 (2004).

Helgadottir, A., Thorleifsson, G., Magnusson, K. P., Gretarsdottir, S., Steinthorsdottir, V., Manolescu, A., Jones, G. T., Rinkel, G. J. E., Blankensteijn, J. D., Ronkainen, A., Jaaskelainen, J. E., Kyo, Y., Lenk, G. M., Sakalihasan, N., Kostulas, K., Gottsater, A., Flex, A., Stefansson, H., Hansen, T., Andersen, G., Weinsheimer, S., Borch-Johnsen, K., Jorgensen, T., Shah, S. H., Quyyumi, A. A., Granger, C. B., Reilly, M. P., Austin, H., Levey, A. I., Vaccarino, V., Palsdottir, E., Walters, G. B., Jonsdottir, T., Snorradottir, S., Magnusdottir, D., Gudmundsson, G., Ferrell, R. E., Sveinbjornsdottir, S., Hernesniemi, J., Niemela, M., Limet, R., Andersen, K., Sigurdsson, G., Benediktsson, R., Verhoeven, E. L. G., Teijink, J. A. W., Grobbee, D. E., Rader, D. J., Collier, D. A., Pedersen, O., Pola, R., Hillert, J., Lindblad, B., Valdimarsson, E. M., Magnadottir, H. B., Wijmenga, C., Tromp, G., Baas, A. F., Ruigrok, Y. M., van Rij, A. M., Kuivaniemi, H., Powell, J. T., Matthiasson, S. E., Gulcher, J. R., Thorgeirsson, G., Kong, A., Thorsteinsdottir, U., & Stefansson, K. The same sequence variant on 9p21 associates with myocardial infarction, abdominal aortic aneurysm and intracranial aneurysm. *Nature Genetics* **40**, 217-224 (2008).

Helgadottir, A., Thorleifsson, G., Manolescu, A., Gretarsdottir, S., Blondal, T., Jonasdottir, A., Jonasdottir, A., Sigurdsson, A., Baker, A., Palsson, A., Masson, G., Gudbjartsson, D. F., Magnusson, K. P., Andersen, K., Levey, A. I., Backman, V. M., Matthiasdottir, S., Jonsdottir, T., Palsson, S., Einarsdottir, H., Gunnarsdottir, S., Gylfason, A., Vaccarino, V., Hooper, W. C., Reilly, M. P., Granger, C. B., Austin, H., Rader, D. J., Shah, S. H., Quyyumi, A. A., Gulcher, J. R., Thorgeirsson, G., Thorsteinsdottir, U., Kong, A., & Stefansson, K. A common variant on chromosome 9p21 affects the risk of myocardial infarction. *Science* **316**, 1491-1493 (2007).

Hellstrom-Lindahl, E., Mousavi, M., Ravid, R., & Nordberg, A. Reduced levels of Abeta 40 and Abeta 42 in brains of smoking controls and Alzheimer's patients. *Neurobiol.Dis.* **15**, 351-360 (2004).

Henzel, W. J., Billeci, T. M., Stults, J. T., Wong, S. C., Grimley, C., & Watanabe, C. Identifying proteins from two-dimensional gels by molecular mass searching of peptide fragments in protein sequence databases. *Proc.Natl.Acad.Sci.U.S A* **90**, 5011-5015 (1993).

Henzel, W. J., Watanabe, C., & Stults, J. T. Protein identification: the origins of peptide mass fingerprinting. *J Am Soc.Mass Spectrom.* **14**, 931-942 (2003).

Heutink, P. & Oostra, B. A. Gene finding in genetically isolated populations. *Human Molecular Genetics* **11**, 2507-2515 (2002).

Hicks, A. A., Petursson, H., Jonsson, T., Stefansson, H., Johannsdottir, H. S., Sainz, J., Frigge, M. L., Kong, A., Gulcher, J. R., Stefansson, K., & Sveinbjornsdottir, S. A susceptibility gene for late-onset idiopathic Parkinson's disease. *Annals of Neurology* **52**, 549-555 (2002).

Higuchi, R., Fockler, C., Dollinger, G., & Watson, R. Kinetic PCR analysis: real-time monitoring of DNA amplification reactions. *Biotechnology (N.Y.)* **11**, 1026-1030 (1993).

HILL, A. B. THE ENVIRONMENT AND DISEASE: ASSOCIATION OR CAUSATION? *Proc.R.Soc.Med.* **58**, 295-300 (1965).

Hill, J. S., Hayden, M. R., Frohlich, J., & Pritchard, P. H. Genetic and environmental factors affecting the incidence of coronary artery disease in heterozygous familial hypercholesterolemia. *Arterioscler.Thromb.* **11**, 290-297 (1991).

Hillenkamp, F., Karas, M., Beavis, R. C., & Chait, B. T. Matrix-Assisted Laser Desorption Ionization Mass-Spectrometry of Biopolymers. *Analytical Chemistry* **63**, A1193-A1202 (1991).

Hipple, J. A., Sommer, H., & Thomas, H. A. A Precise Method of Determining the Faraday by Magnetic Resonance. *Physical Review* **76**, 1877-1878 (1949).

Hirschhorn, J. N. & Daly, M. J. Genome-wide association studies for common diseases and complex traits. *Nature Reviews Genetics* **6**, 95-108 (2005).

Hoglund, K., Thelen, K. M., Syversen, S., Sjogren, M., von Bergmann, K., Wallin, A., Vanmechelen, E., Vanderstichele, H., Lutjohann, D., & Blennow, K. The effect of simvastatin treatment on the amyloid precursor protein and brain cholesterol metabolism in patients with Alzheimer's disease. *Dementia and Geriatric Cognitive Disorders* **19**, 256-265 (2005).

Holcomb, L., Gordon, M. N., McGowan, E., Yu, X., Benkovic, S., Jantzen, P., Wright, K., Saad, I., Mueller, R., Morgan, D., Sanders, S., Zehr, C., O'Campo, K., Hardy, J., Prada, C. M., Eckman, C., Younkin, S., Hsiao, K., & Duff, K. Accelerated Alzheimer-type phenotype in transgenic mice carrying both mutant amyloid precursor protein and presenilin 1 transgenes. *Nat.Med.* **4**, 97-100 (1998).

Hollyfield, J. G., Salomon, R. G., & Crabb, J. W. Proteomic approaches to understanding age-related macular degeneration. *Retinal Degenerations: Mechanisms and Experimental Therapy* **533**, 83-89 (2003).

Hope, G. M., Dawson, W. W., Engel, H. M., Ulshafer, R. J., Kessler, M. J., & Sherwood, M. B. A primate model for age related macular drusen. *Br.J Ophthalmol.* **76**, 11-16 (1992).

Hopper, J. L. & Mathews, J. D. Extensions to Multivariate Normal-Models for Pedigree Analysis. *Annals of Human Genetics* **46**, 373-383 (1982).

Horton R. L. (1978). The General Linear Model. New York: McGraw-Hill, Inc.

Hsiao, K., Chapman, P., Nilsen, S., Eckman, C., Harigaya, Y., Younkin, S., Yang, F., & Cole, G. Correlative memory deficits, Abeta elevation, and amyloid plaques in transgenic mice. *Science* **274**, 99-102 (1996).

Hsu, F. C., Zaccaro, D. J., Lange, L. A., Arnett, D. K., Langefeld, C. D., Wagenknecht, L. E., Herrington, D. M., Beck, S. R., Freedman, B. I., Bowden, D. W., & Rich, S. S. The impact of pedigree structure on heritability estimates for pulse pressure in three studies. *Hum.Hered.* **60**, 63-72 (2005).

Huang, H. L., Stasyk, T., Morandell, S., Mogg, M., Schreiber, M., Feuerstein, I., Huck, C. W., Stecher, G., Bonn, G. K., & Huber, L. A. Enrichment of low-abundant serum proteins by albumin/immunoglobulin G immunoaffinity depletion under partly denaturing conditions. *Electrophoresis* **26**, 2843-2849 (2005).

Huentelman, M. J., Papassotiropoulos, A., Craig, D. W., Hoerndli, F. J., Pearson, J. V., Huynh, K. D., Corneveaux, J., Hanggi, J., Mondadori, C. R., Buchmann, A., Reiman, E. M., Henke, K., de Quervain, D. J., & Stephan, D. A. Calmodulin-binding transcription activator 1 (CAMTA1) alleles predispose human episodic memory performance. *Hum.Mol.Genet.* **16**, 1469-1477 (2007).

Hughes, A. E., Orr, N., Esfandiary, H., Diaz-Torres, M., Goodship, T., & Chakravarthy, U. A common CFH haplotype, with deletion of CFHR1 and CFHR3, is associated with lower risk of age-related macular degeneration. *Nature Genetics* **38**, 1173-1177 (2006).

Hughes, A. E., Orr, N., Patterson, C., Esfandiary, H., Hogg, R., McConnell, V., Silvestri, G., & Chakravarthy, U. Neovascular age-related macular degeneration risk based on CFH, LOC387715/HTRA1, and smoking. *PLoS Med.* **4**, e355 (2007).

Hutchens, T. W. & Yip, T. T. New Desorption Strategies for the Mass-Spectrometric Analysis of Macromolecules. *Rapid Communications in Mass Spectrometry* **7**, 576-580 (1993).

Iglewicz B. & Hoaglin D. C. (1993). How to detect and Handle Outliers. ASQ Quality Press.

Imamura, Y., Noda, S., Hashizume, K., Shinoda, K., Yamaguchi, M., Uchiyama, S., Shimizu, T., Mizushima, Y., Shirasawa, T., & Tsubota, K. Drusen, choroidal neovascularization, and retinal pigment epithelium dysfunction in SOD1-deficient mice: a model of age-related macular degeneration. *Proc.Natl.Acad.Sci.U.S A* **103**, 11282-11287 (2006).

Irizarry, M. C., Soriano, F., McNamara, M., Page, K. J., Schenk, D., Games, D., & Hyman, B. T. Abeta deposition is associated with neuropil changes, but not with overt neuronal loss in the human amyloid precursor protein V717F (PDAPP) transgenic mouse. *J Neurosci.* **17**, 7053-7059 (1997).

Isaacs, A., Sayed-Tabatabaei, F. A., Aulchenko, Y. S., Zillikens, M. C., Sijbrands, E. J., Schut, A. F., Rutten, W. P., Pols, H. A., Witteman, J. C., Oostra, B. A., & van Duijn, C. M. Heritabilities, apolipoprotein E, and effects of inbreeding on plasma lipids in a genetically isolated population: the Erasmus Rucphen Family Study. *Eur.J Epidemiol.* **22**, 99-105 (2007).

Ishida, B. Y., Bailey, K. R., Duncan, K. G., Chalkley, R. J., Burlingame, A. L., Kane, J. P., & Schwartz, D. M. Regulated expression of apolipoprotein E by human retinal pigment epithelial cells. *J Lipid Res* **45**, 263-271 (2004).

Ivkovic, V., Vitart, V., Rudan, I., Janicijevic, B., Smolej-Narancic, N., Skaric-Juric, T., Barbalic, M., Polasek, O., Kolcic, I., Biloglav, Z., Visscher, P. M., Hayward, C., Hastie, N. D., Anderson, N., Campbell, H., Wright, A. F., Rudan, P., & Deary, I. J. The Eysenck personality factors: Psychometric structure, reliability, heritability and phenotypic and genetic correlations with psychological distress in an isolated Croatian population. *Personality and Individual Differences* **42**, 123-133 (2007).

Iyengar, S. K., Song, D., Klein, B. E., Klein, R., Schick, J. H., Humphrey, J., Millard, C., Liptak, R., Russo, K., Jun, G., Lee, K. E., Fijal, B., & Elston, R. C. Dissection of genomewide-scan data in extended families reveals a major locus and oligogenic susceptibility for age-related macular degeneration. *Am J Hum.Genet.* **74**, 20-39 (2004).

Jakobsdottir, J., Conley, Y. P., Weeks, D. E., Mah, T. S., Ferrell, R. E., & Gorin, M. B. Susceptibility genes for age-related maculopathy on chromosome 10q26. *American Journal of Human Genetics* **77**, 389-407 (2005).

Jarva, H., Jokiranta, T. S., Hellwage, J., Zipfel, P. F., & Meri, S. Regulation of complement activation by C-reactive protein: targeting the complement inhibitory activity of factor H by an interaction with short consensus repeat domains 7 and 8-11. *J Immunol.* **163**, 3957-3962 (1999).

Jeran N. & Havas D. (2007). Mitochondrial DNA Haplogroup Diversity of the Island of Vis (Poster). 5th ISABS Conference in Forensic Genetics and Molecular Anthropology. Split, Hrvatska.

Johnson, L. V., Leitner, W. P., Staples, M. K., & Anderson, D. H. Complement activation and inflammatory processes in Drusen formation and age related macular degeneration. *Exp.Eye Res* **73**, 887-896 (2001).

Kaessmann, H., Zollner, S., Gustafsson, A. C., Wiebe, V., Laan, M., Lundeberg, J., Uhlen, M., & Paabo, S. Extensive linkage disequilibrium in small human populations in Eurasia. *American Journal of Human Genetics* **70**, 673-685 (2002).

Kaiser, T., Wittke, S., Just, I., Krebs, R., Bartel, S., Fliser, D., Mischak, H., & Weissinger, E. M. Capillary electrophoresis coupled to mass spectrometer for automated and robust polypeptide determination in body fluids for clinical use. *Electrophoresis* **25**, 2044-2055 (2004).

Kanai, M., Matsubara, E., Isoe, K., Urakami, K., Nakashima, K., Arai, H., Sasaki, H., Abe, K., Iwatsubo, T., Kosaka, T., Watanabe, M., Tomidokoro, Y., Shizuka, M., Mizushima, K., Nakamura, T., Igeta, Y., Ikeda, Y., Amari, M., Kawarabayashi, T., Ishiguro, K., Harigaya, Y., Wakabayashi, K., Okamoto, K., Hirai, S., & Shoji, M. Longitudinal study of cerebrospinal fluid levels of tau, A beta1-40, and A beta1-42(43) in Alzheimer's disease: a study in Japan. *Ann.Neurol.* **44**, 17-26 (1998).

Kanda, A., Chen, W., Othman, M., Branham, K. E. H., Brooks, M., Khanna, R., He, S., Lyons, R., Abecasis, G. R., & Swaroop, A. A variant of mitochondrial protein LOC387715/ARMS2, not HTRA1, is strongly associated with age-related macular degeneration. *Proceedings of the National Academy of Sciences of the United States of America* **104**, 16227-16232 (2007).

Kang, J. E., Logroscino, G., De Vivo, I., Hunter, D., & Grodstein, F. Apolipoprotein E, cardiovascular disease and cognitive function in aging women. *Neurobiology of Aging* **26**, 475-484 (2005).

Karan, G., Lillo, C., Yang, Z., Cameron, D. J., Locke, K. G., Zhao, Y., Thirumalaichary, S., Li, C., Birch, D. G., Vollmer-Snarr, H. R., Williams, D. S., & Zhang, K. Lipofuscin accumulation, abnormal electrophysiology, and photoreceptor degeneration in mutant ELOVL4 transgenic mice: a model for macular degeneration. *Proc.Natl.Acad.Sci.U.S A* **102**, 4164-4169 (2005).

Karas, M. & Hillenkamp, F. Laser Desorption Ionization of Proteins with Molecular Masses Exceeding 10000 Daltons. *Analytical Chemistry* **60**, 2299-2301 (1988).

Kemper, C., Gigli, I., & Zipfel, P. F. Conservation of plasma regulatory proteins of the complement system in evolution: humans and fish. *Exp.Clin.Immunogenet.* **17**, 55-62 (2000).

Kemper, C., Zipfel, P. F., & Gigli, I. The complement cofactor protein (SBP1) from the barred sand bass (Paralabrax nebulifer) mediates overlapping regulatory activities of both human C4b binding protein and factor H. *J Biol.Chem.* **273**, 19398-19404 (1998).

Kenealy, S. J., Schmidt, S., Agarwal, A., Postel, E. A., De La Paz, M. A., Pericak-Vance, M. A., & Haines, J. L. Linkage analysis for age-related macular degeneration supports a gene on chromosome 10q26. *Mol.Vis.* **10**, 57-61 (2004).

King, D. C., Taylor, J., Elnitski, L., Chiaromonte, F., Miller, W., & Hardison, R. C. Evaluation of regulatory potential and conservation scores for detecting cis-regulatory modules in aligned mammalian genome sequences. *Genome Res* **15**, 1051-1060 (2005).

Klaric, I. M., Barac, L., Bukovic, D., Furac, I., Geber, G., Janicijevic, B., Kubat, M., Pericic, M., Pupic, B. V., & Rudan, P. Short tandem repeat (STR) variation in eight village populations of the island of Korcula (Croatia). *Ann.Hum.Biol.* **28**, 281-294 (2001).

Klaver, C. C. W., Kliffen, M., van Duijn, C. M., Hofman, A., Cruts, M., Grobbee, D. E., van Broeckhoven, C., & De Jong, P. T. V. M. Genetic association of apolipoprotein E with age-related macular degeneration. *American Journal of Human Genetics* **63**, 200-206 (1998a).

Klaver, C. C., Wolfs, R. C., Assink, J. J., van Duijn, C. M., Hofman, A., & de Jong, P. T. Genetic risk of age-related maculopathy. Population-based familial aggregation study. *Arch.Ophthalmol.* **116**, 1646-1651 (1998b).

Klein, B. E., Klein, R., Lee, K. E., Moore, E. L., & Danforth, L. Risk of incident age-related eye diseases in people with an affected sibling : The Beaver Dam Eye Study. *Am J Epidemiol.* **154**, 207-211 (2001).

Klein, M. L., Schultz, D. W., Edwards, A., Matise, T. C., Rust, K., Berselli, C. B., Trzupek, K., Weleber, R. G., Ott, J., Wirtz, M. K., & Acott, T. S. Age-related macular degeneration. Clinical features in a large family and linkage to chromosome 1q. *Arch.Ophthalmol.* **116**, 1082-1088 (1998).

Klein, R. J. Power analysis for genome-wide association studies. *BMC Genet.* **8**, 58 (2007).

Klein, R. J., Zeiss, C., Chew, E. Y., Tsai, J. Y., Sackler, R. S., Haynes, C., Henning, A. K., SanGiovanni, J. P., Mane, S. M., Mayne, S. T., Bracken, M. B., Ferris, F. L., Ott, J., Barnstable, C., & Hoh, J. Complement factor H polymorphism in age-related macular degeneration. *Science* **308**, 385-389 (2005).

Klein, R., Klein, B. E., & Cruickshanks, K. J. The prevalence of age-related maculopathy by geographic region and ethnicity. *Prog.Retin.Eye Res* **18**, 371-389 (1999).

Klein, R., Klein, B. E., & Linton, K. L. Prevalence of age-related maculopathy. The Beaver Dam Eye Study. *Ophthalmology* **99**, 933-943 (1992).

Klein, R., Klein, B. E., Jensen, S. C., & Meuer, S. M. The five-year incidence and progression of age-related maculopathy: the Beaver Dam Eye Study. *Ophthalmology* **104**, 7-21 (1997).

Kleinjan, D. A., Bancewicz, R. M., Gautier, P., Dahm, R., Schonthaler, H. B., Damante, G., Seawright, A., Hever, A. M., Yeyati, P. L., van, H., V, & Coutinho, P. Subfunctionalization of Duplicated Zebrafish pax6 Genes by cis-Regulatory Divergence. *PLoS Genet.* **4**, e29 (2008).

Kolbe, D., Taylor, J., Elnitski, L., Eswara, P., Li, J., Miller, W., Hardison, R., & Chiaromonte, F. Regulatory potential scores from genome-wide three-way alignments of human, mouse, and rat. *Genome Res* **14**, 700-707 (2004).

Kondo, N., Honda, S., Ishibashi, K., Tsukahara, Y., & Negi, A. LOC387715/HTRA1 variants in polypoidal choroidal vasculopathy and age-related macular degeneration in a Japanese population. *Am J Ophthalmol.* **144**, 608-612 (2007).

Koomen, J. M., Shill, L. N., Coombes, K. R., Li, D. H., Xiao, L. C., Fidler, I. J., Abbruzzese, J. L., & Kobayashi, R. Plasma protein profiling for diagnosis of pancreatic cancer reveals the presence of host response proteins. *Clinical Cancer Research* **11**, 1110-1118 (2005).

Kruglyak, L. What is significant in whole-genome linkage disequilibrium studies? *Am J Hum.Genet.* **61**, 810-812 (1997).

Kuo, Y. M., Emmerling, M. R., Bisgaier, C. L., Essenburg, A. D., Lampert, H. C., Drumm, D., & Roher, A. E. Elevated low-density lipoprotein in Alzheimer's disease correlates with brain abeta 1-42 levels. *Biochem.Biophys.Res Commun.* **252**, 711-715 (1998).

Kuo, Y. M., Emmerling, M. R., Lampert, H. C., Hempelman, S. R., Kokjohn, T. A., Woods, A. S., Cotter, R. J., & Roher, A. E. High levels of circulating A beta 42 are sequestered by plasma proteins in Alzheimer's disease. *Biochemical and Biophysical Research Communications* **257**, 787-791 (1999).

Kuo, Y. M., Kokjohn, T. A., Kalback, W., Luehrs, D., Galasko, D. R., Chevallier, N., Koo, E. H., Emmerling, M. R., & Roher, A. E. Amyloid-beta peptides interact with plasma proteins and erythrocytes: Implications for their quantitation in plasma. *Biochemical and Biophysical Research Communications* **268**, 750-756 (2000).

Kuo, Y. M., Kokjohn, T. A., Watson, M. D., Woods, A. S., Cotter, R. J., Sue, L. I., Kalback, W. M., Emmerling, M. R., Beach, T. G., & Roher, A. E. Elevated A beta 42 in skeletal muscle of Alzheimer disease patients suggests peripheral alterations of A beta PP metabolism. *American Journal of Pathology* **156**, 797-805 (2000).

Laan, M. & Paabo, S. Demographic history and linkage disequilibrium in human populations. *Nat.Genet.* **17**, 435-438 (1997).

Laitinen, T., Polvi, A., Rydman, P., Vendelin, J., Pulkkinen, V., Salmikangas, P., Makela, S., Rehn, M., Pirskanen, A., Rautanen, A., Zucchelli, M., Gullsten, H., Leino, M., Alenius, H., Petays, T., Haahtela, T., Laitinen, A., Laprise, C., Hudson, T. J., Laitinen, L. A., & Kere, J. Characterization of a common susceptibility locus for asthma-related traits. *Science* **304**, 300-304 (2004).

Lander, E. & Kruglyak, L. Genetic dissection of complex traits: guidelines for interpreting and reporting linkage results. *Nat.Genet.* **11**, 241-247 (1995).

Lander, E. S. & Schork, N. J. Genetic dissection of complex traits. *Science* **265**, 2037-2048 (1994).

Lau, L. I., Chen, S. J., Cheng, C. Y., Yen, M. Y., Lee, F. L., Lin, M. W., Hsu, W. M., & Wei, Y. H. Association of the Y402H polymorphism in complement factor H gene and neovascular age-related macular degeneration in Chinese patients. *Invest Ophthalmol.Vis.Sci.* **47**, 3242-3246 (2006).

Lee, J. Y., Cole, T. B., Palmiter, R. D., Suh, S. W., & Koh, J. Y. Contribution by synaptic zinc to the gender-disparate plaque formation in human Swedish mutant APP transgenic mice. *Proc.Natl.Acad.Sci.U.S A* **99**, 7705-7710 (2002).

Levy, M., Halbwachs-Mecarelli, L., Gubler, M. C., Kohout, G., Bensenouci, A., Niaudet, P., Hauptmann, G., & Lesavre, P. H deficiency in two brothers with atypical dense intramembranous deposit disease. *Kidney Int.* **30**, 949-956 (1986).

Lewczuk, P., Esselmann, H., Otto, M., Maler, J. M., Henkel, A. W., Henkel, M. K., Eikenberg, O., Antz, C., Krause, W. R., Reulbach, U., Kornhuber, J., & Wiltfang, J. Neurochemical diagnosis of Alzheimer's dementia by CSF A beta 42, A beta 42/A beta 40 ratio and total tau. *Neurobiology of Aging* **25**, 273-281 (2004).

Li, C. M., Clark, M. E., Chimento, M. F., & Curcio, C. A. Apolipoprotein localization in isolated drusen and retinal apolipoprotein gene expression. *Invest Ophthalmol.Vis.Sci.* **47** , 3119-3128 (2006a).

Li, H., Wetten, S., Li, L., St Jean, P. L., Upmanyu, R., Surh, L., Hosford, D., Barnes, M. R., Briley, J. D., Borrie, M., Coletta, N., Delisle, R., Dhalla, D., Ehm, M. G., Feldman, H. H., Fornazzari, L., Gauthier, S., Goodgame, N., Guzman, D., Hammond, S., Hollingworth, P., Hsiung, G. Y., Johnson, J., Kelly, D. D., Keren, R., Kertesz, A., King, K. S., Lovestone, S., Loy-English, I., Matthews, P. M., Owen, M. J., Plumpton, M., Pryse-Phillips, W., Prinjha, R. K., Richardson, J. C., Saunders, A., Slater, A. J., George-Hyslop, P. H., Stinnett, S. W., Swartz, J. E., Taylor, R. L., Wherrett, J., Williams, J., Yarnall, D. P., Gibson, R. A., Irizarry, M. C., Middleton, L. T., & Roses, A. D. Candidate single-nucleotide polymorphisms from a genomewide association study of Alzheimer disease. *Arch.Neurol.* **65**, 45-53 (2008).

Li, M., Atmaca-Sonmez, P., Othman, M., Branham, K. E. H., Khanna, R., Wade, M. S., Li, Y., Liang, L. M., Zareparsi, S., Swaroop, A., & Abecasis, G. R. CFH haplotypes without the Y402H coding variant show strong association with susceptibility to age-related macular degeneration. *Nature Genetics* **38**, 1049-1054 (2006).

Li, Y., Xu, L., Jonas, J. B., Yang, H., Ma, Y., & Li, J. Prevalence of age-related maculopathy in the adult population in China: the Beijing eye study. *Am J Ophthalmol.* **142**, 788-793 (2006b).

Lindahl, G., Sjobring, U., & Johnsson, E. Human complement regulators: a major target for pathogenic microorganisms. *Curr.Opin.Immunol.* **12**, 44-51 (2000).

Lindenbaum, R. H., Clarke, G., Patel, C., Moncrieff, M., & Hughes, J. T. Muscular dystrophy in an X; 1 translocation female suggests that Duchenne locus is on X chromosome short arm. *J Med.Genet.* **16**, 389-392 (1979).

Little, S. P., Dixon, E. P., Norris, F., Buckley, W., Becker, G. W., Johnson, M., Dobbins, J. R., Wyrick, T., Miller, J. R., MacKellar, W., Hepburn, D., Corvalan, J., McClure, D., Liu, X., Stephenson, D., Clemens, J., & Johnstone, E. M. Zyme, a novel and potentially amyloidogenic enzyme cDNA isolated from Alzheimer's disease brain. *J Biol.Chem.* **272**, 25135-25142 (1997).

Liu, X. Z., Hope, C., Liang, C. Y., Zou, J. M., Xu, L. R., Cole, T., Mueller, R. F., Bundey, S., Nance, W., Steel, K. P., & Brown, S. D. A mutation (2314delG) in the Usher syndrome type IIA gene: high prevalence and phenotypic variation. *Am J Hum.Genet.* **64**, 1221-1225 (1999).

Lopezlarrea, C., Dieguez, M. A., Enguix, A., Dominguez, O., Marin, B., & Gomez, E. A Familial Deficiency of Complement Factor-H. *Biochemical Society Transactions* **15**, 648-649 (1987).

Lotery, A. & Trump, D. Progress in defining the molecular biology of age related macular degeneration. *Hum.Genet.* **122**, 219-236 (2007).

Lucotte, G., Loirat, F., & Hazout, S. Pattern of gradient of apolipoprotein E allele *4 frequencies in western Europe. *Hum.Biol.* **69**, 253-262 (1997).

Lucotte, G., Visvikis, S., Leininger-Muler, B., David, F., Berriche, S., Reveilleau, S., Couderc, R., Babron, M. C., Aguillon, D., & Siest, G. Association of apolipoprotein E allele epsilon 4 with late-onset sporadic Alzheimer's disease. *Am J Med.Genet.* **54**, 286-288 (1994).

Mackay, T. F. C. Quantitative trait loci in Drosophila. *Nature Reviews Genetics* **2**, 11-20 (2001).

Mahley, R. W. & Huang, Y. D. Apolipoprotein E: from atherosclerosis to Alzheimer's disease and beyond. *Current Opinion in Lipidology* **10**, 207-217 (1999).

Mahley, R. W. & Rall, S. C. Apolipoprotein E: Far more than a lipid transport protein. *Annual Review of Genomics and Human Genetics* **1**, 507-537 (2000).

Mahley, R. W. Apolipoprotein-e - Cholesterol Transport Protein with Expanding Role in Cell Biology. *Science* **240**, 622-630 (1988).

Majewski, J., Schultz, D. W., Weleber, R. G., Schain, M. B., Edwards, A. O., Matise, T. C., Acott, T. S., Ott, J., & Klein, M. L. Age-related macular degeneration--a genome scan in extended families. *Am J Hum.Genet.* **73**, 540-550 (2003).

Malek, G., Johnson, L. V., Mace, B. E., Saloupis, P., Schmechel, D. E., Rickman, D. W., Toth, C. A., Sullivan, P. M., & Bowes, R. C. Apolipoprotein E allele-dependent pathogenesis: a model for age-related retinal degeneration. *Proc.Natl.Acad.Sci.U.S A* **102**, 11900-11905 (2005).

Maller, J. B., Fagerness, J. A., Reynolds, R. C., Neale, B. M., Daly, M. J., & Seddon, J. M. Variation in complement factor 3 is associated with risk of age-related macular degeneration. *Nat.Genet.* **39**, 1200-1201 (2007).

Maller, J., George, S., Purcell, S., Fagerness, J., Altshuler, D., Daly, M. J., & Seddon, J. M. Common variation in three genes, including a noncoding variant in CFH, strongly influences risk of age-related macular degeneration. *Nature Genetics* **38**, 1055-1059 (2006).

Maltzman, B. A., Mulvihill, M. N., & Greenbaum, A. Senile macular degeneration and risk factors: a case-control study. *Ann.Ophthalmol.* **11**, 1197-1201 (1979).

Mann, J. I., Lewis, B., Shepherd, J., Winder, A. F., Fenster, S., Rose, L., & Morgan, B. Blood lipid concentrations and other cardiovascular risk factors: distribution, prevalence, and detection in Britain. *Br.Med.J (Clin.Res Ed)* **296**, 1702-1706 (1988).

Mann, K. M., Thorngate, F. E., Katoh-Fukui, Y., Hamanaka, H., Williams, D. L., Fujita, S., & Lamb, B. T. Independent effects of APOE on cholesterol metabolism and brain A beta levels in an Alzheimer disease mouse model. *Human Molecular Genetics* **13**, 1959-1968 (2004).

Mares-Perlman, J. A., Brady, W. E., Klein, R., VandenLangenberg, G. M., Klein, B. E., & Palta, M. Dietary fat and age-related maculopathy. *Arch.Ophthalmol.* **113**, 743-748 (1995).

Masters, C. L., Simms, G., Weinman, N. A., Multhaup, G., Mcdonald, B. L., & Beyreuther, K. Amyloid Plaque Core Protein in Alzheimer-Disease and Down Syndrome.

Proceedings of the National Academy of Sciences of the United States of America **82**, 4245-4249 (1985).

Matsumoto, A., Minami, M., & Matsumoto, R. The beta-amyloid epitope masking activity in human brain is identified as albumin. *Neuroreport* **8**, 3297-3301 (1997).

Matthews, K. S. DNA looping. *Microbiol.Rev.* **56**, 123-136 (1992).

Mayeux, R. Mapping the new frontier: complex genetic disorders. *J Clin.Invest* **115**, 1404-1407 (2005).

Mayeux, R., Honig, L. S., Tang, M. X., Manly, J., Stern, Y., Schupf, N., & Mehta, P. D. Plasma A beta 40 and A beta 42 and Alzheimer's disease - Relation to age, mortality, and risk. *Neurology* **61**, 1185-1190 (2003).

Mayeux, R., Tang, M. X., Jacobs, D. M., Manly, J., Bell, K., Merchant, C., Small, S. A., Stern, Y., Wisniewski, H. M., & Mehta, P. D. Plasma amyloid beta-peptide 1-42 and incipient Alzheimer's disease. *Annals of Neurology* **46**, 412-416 (1999).

Mclafferty, F. W. Tandem Mass-Spectrometry (Ms-Ms). *Abstracts of Papers of the American Chemical Society* **182**, 75-ANYL (1981).

Mehta, P. D., Pirttila, T., Mehta, S. P., Sersen, E. A., Aisen, P. S., & Wisniewski, H. M. Plasma and cerebrospinal fluid levels of amyloid beta proteins 1-40 and 1-42 in Alzheimer disease. *Arch.Neurol.* **57**, 100-105 (2000).

Mehta, P. D., Pirttila, T., Patrick, B. A., Barshatzky, M., & Mehta, S. P. Amyloid beta protein 1-40 and 1-42 levels in matched cerebrospinal fluid and plasma from patients with Alzheimer disease. *Neuroscience Letters* **304**, 102-106 (2001).

Menard S. (1995). Applied Logistic Regression Analysis. SAGE University paper series on quantitative applications in the social sciences. Thousand Oaks, CA.

Meri, S. & Pangburn, M. K. Discrimination between activators and nonactivators of the alternative pathway of complement: regulation via a sialic acid/polyanion binding site on factor H. *Proc.Natl.Acad.Sci.U.S A* **87**, 3982-3986 (1990).

Merrell, K., Southwick, K., Graves, S. W., Esplin, M. S., Lewis, N. E., & Thulin, C. D. Analysis of low-abundance, low-molecular-weight serum proteins using mass spectrometry. *J Biomol.Tech.* **15**, 238-248 (2004).

Merril, C. R. & Harrington, M. G. "Ultrasensitive" silver stains: their use exemplified in the study of normal human cerebrospinal fluid proteins separated by two-dimensional electrophoresis. *Clin.Chem.* **30**, 1938-1942 (1984).

Miller R. G. (1981). Simultaneous Statistical Inference. New York: McGraw-Hill.

Misasi, R., Huemer, H. P., Schwaeble, W., Solder, E., Larcher, C., & Dierich, M. P. Human complement factor H: an additional gene product of 43 kDa isolated from human plasma shows cofactor activity for the cleavage of the third component of complement. *Eur.J Immunol.* **19**, 1765-1768 (1989).

Mitsui, S., Okui, A., Uemura, H., Mizuno, T., Yamada, T., Yamamura, Y., & Yamaguchi, N. Decreased cerebrospinal fluid levels of neurosin (KLK6), an aging-related protease, as a possible new risk factor for Alzheimer's disease. *Ann.N.Y.Acad.Sci.* **977**, 216-223 (2002).

Morimatsu, M., Hirai, S., Muramatsu, A., & Yoshikawa, M. Senile degenerative brain lesions and dementia. *J Am Geriatr.Soc.* **23**, 390-406 (1975).

Morton, N. E. Sequential Tests for the Detection of Linkage. *American Journal of Human Genetics* **7**, 277-318 (1955).

Mullins, R. F., Russell, S. R., Anderson, D. H., & Hageman, G. S. Drusen associated with aging and age-related macular degeneration contain proteins common to extracellular deposits associated with atherosclerosis, elastosis, amyloidosis, and dense deposit disease. *FASEB J* **14**, 835-846 (2000).

Myers R. (1990). Classical and Modern Regression with Applications (2nd Edition). Boston, MA.

Myers, A. J., Marshall, H., Holmans, P., Compton, D., Crook, R. J. P., Mander, A. P., Nowotny, P., Smemo, S., Dunstan, M., Jehu, L., Wang, J. C., Hamshere, M., Morris, J. C., Norton, J., Chakraventy, S., Tunstall, N., Lovestone, S., Petersen, R., O'Donovan, M., Jones, L., Williams, J., Owen, M. J., John, H. Y., & Goate, A. Variation in the urokinase-plasminogen activator gene does not explain the chromosome 10 linkage signal for late onset AD. *American Journal of Medical Genetics Part B-Neuropsychiatric Genetics* **124B**, 29-37 (2004).

Myers, R. H., Schaefer, E. J., Wilson, P. W., D'Agostino, R., Ordovas, J. M., Espino, A., Au, R., White, R. F., Knoefel, J. E., Cobb, J. L., McNulty, K. A., Beiser, A., & Wolf, P. A. Apolipoprotein E epsilon4 association with dementia in a population-based study: The Framingham study. *Neurology* **46**, 673-677 (1996).

Newman, D. L., Abney, M., Dytch, H., Parry, R., McPeek, M. S., & Ober, C. Major loci influencing serum triglyceride levels on 2q14 and 9p21 localized by homozygosity-by-descent mapping in a large Hutterite pedigree. *Hum.Mol.Genet.* **12**, 137-144 (2003).

Nicchia, G. P., Rossi, A., Nudel, U., Svelto, M., & Frigeri, A. Dystrophin-dependent and -independent AQP4 pools are expressed in the mouse brain. *Glia* (2008).

Nozaki, M., Raisler, B. J., Sakurai, E., Sarma, J. V., Barnum, S. R., Lambris, J. D., Chen, Y., Zhang, K., Ambati, B. K., Baffi, J. Z., & Ambati, J. Drusen complement components C3a and C5a promote choroidal neovascularization. *Proc.Natl.Acad.Sci.U.S A* **103**, 2328-2333 (2006).

Nussinov, R. The eukaryotic CCAAT and TATA boxes, DNA spacer flexibility and looping. *J Theor.Biol.* **155**, 243-270 (1992).

Oakley, B. R., Kirsch, D. R., & Morris, N. R. A simplified ultrasensitive silver stain for detecting proteins in polyacrylamide gels. *Anal.Biochem.* **105**, 361-363 (1980).

Oddo, S., Billings, L., Kesslak, J. P., Cribbs, D. H., & LaFerla, F. M. Abeta immunotherapy leads to clearance of early, but not late, hyperphosphorylated tau aggregates via the proteasome. *Neuron* **43**, 321-332 (2004).

Oddo, S., Caccamo, A., Shepherd, J. D., Murphy, M. P., Golde, T. E., Kayed, R., Metherate, R., Mattson, M. P., Akbari, Y., & LaFerla, F. M. Triple-transgenic model of Alzheimer's disease with plaques and tangles: intracellular Abeta and synaptic dysfunction. *Neuron* **39**, 409-421 (2003).

Ogata, K., Sato, K., & Tahirov, T. H. Eukaryotic transcriptional regulatory complexes: cooperativity from near and afar. *Curr.Opin.Struct.Biol.* **13**, 40-48 (2003).

Okamoto, H., Umeda, S., Obazawa, M., Minami, M., Noda, T., Mizota, A., Honda, M., Tanaka, M., Koyama, R., Takagi, I., Sakamoto, Y., Saito, Y., Miyake, Y., & Iwata, T. Complement factor H polymorphisms in Japanese population with age-related macular degeneration. *Mol.Vis.* **12**, 156-158 (2006).

Ong, S. E., Blagoev, B., Kratchmarova, I., Kristensen, D. B., Steen, H., Pandey, A., & Mann, M. Stable isotope labeling by amino acids in cell culture, SILAC, as a simple and accurate approach to expression proteomics. *Molecular & Cellular Proteomics* **1**, 376-386 (2002).

Ono, K., Hasegawa, K., Yamada, M., & Naiki, H. Nicotine breaks down preformed Alzheimer's beta-amyloid fibrils in vitro. *Biol.Psychiatry* **52**, 880-886 (2002).

Oshima, Y., Ishibashi, T., Murata, T., Tahara, Y., Kiyohara, Y., & Kubota, T. Prevalence of age related maculopathy in a representative Japanese population: the Hisayama study. *Br.J Ophthalmol.* **85**, 1153-1157 (2001).

Ott J. (1991). Analysis of Human Genetic Linkage (Rev. Ed.). Johns Hopkins University Press, Baltimore.

Pajukanta, P., Lilja, H. E., Sinsheimer, J. S., Cantor, R. M., Lusis, A. J., Gentile, M., Duan, X. Q. J., Soro-Paavonen, A., Naukkarinen, J., Saarela, J., Laakso, M., Ehnholm, C., Taskinen, M. R., & Peltonen, L. Familial combined hyperlipidemia is associated with upstream transcription factor 1 (USF1). *Nature Genetics* **36**, 371-376 (2004).

Pajukanta, P., Nuotio, I., Terwilliger, J. D., Porkka, K. V. K., Viikari, J. S. A., Laakso, M., Taskinen, M. R., Ehnholm, C., & Peltonen, L. Linkage of familial combined hyperlipidaemia (FCHL) to chromosome 1q21-q23: a new locus predisposing to premature coronary heart disease. *Atherosclerosis* **138**, S3 (1998).

Pardo, L. M., MacKay, I., Oostra, B., van Duijn, C. M., & Aulchenko, Y. S. The effect of genetic drift in a young genetically isolated population. *Annals of Human Genetics* **69**, 288-295 (2005).

Patton, E. E. & Zon, L. I. Taking Human Cancer Genes to the Fish: A Transgenic Model of Melanoma in Zebrafish. *Zebrafish.* **1**, 363-368 (2005).

Pepys, M. B. C-reactive protein is neither a marker nor a mediator of atherosclerosis. *Nat.Clin.Pract.Nephrol.* (2008).

Pericak-Vance, M. A., Bebout, J. L., Gaskell, P. C., Jr., Yamaoka, L. H., Hung, W. Y., Alberts, M. J., Walker, A. P., Bartlett, R. J., Haynes, C. A., Welsh, K. A., & . Linkage studies in familial Alzheimer disease: evidence for chromosome 19 linkage. *Am J Hum.Genet.* **48**, 1034-1050 (1991).

Perkins, D. N., Pappin, D. J. C., Creasy, D. M., & Cottrell, J. S. Probability-based protein identification by searching sequence databases using mass spectrometry data. *Electrophoresis* **20**, 3551-3567 (1999).

Perry, E. K., Court JA, Lloyd, S., Johnson, M., Griffiths, M. H., Spurden, D., Piggott, M. A., Turner, J., & Perry, R. H. Beta-amyloidosis in normal aging and transmitter signaling in human temporal lobe. *Ann.N.Y.Acad.Sci.* **777**, 388-392 (1996).

Perry, E., Martin-Ruiz, C., Lee, M., Griffiths, M., Johnson, M., Piggott, M., Haroutunian, V., Buxbaum, J. D., Nasland, J., Davis, K., Gotti, C., Clementi, F., Tzartos, S., Cohen, O., Soreq, H., Jaros, E., Perry, R., Ballard, C., McKeith, I., & Court, J. Nicotinic receptor subtypes in human brain ageing, Alzheimer and Lewy body diseases. *Eur.J Pharmacol.* **393**, 215-222 (2000).

Pieper, R., Su, Q., Gatlin, C. L., Huang, S. T., Anderson, N. L., & Steiner, S. Multi-component immunoaffinity subtraction chromatography: An innovative step towards a comprehensive survey of the human plasma proteome. *Proteomics* **3**, 422-432 (2003).

Pilia, G., Chen, W. M., Scuteri, A., Orru, M., Albai, G., Dei, M., Lai, S., Usala, G., Lai, M., Loi, P., Mameli, C., Vacca, L., Deiana, M., Olla, N., Masala, M., Cao, A., Najjar, S. S., Terracciano, A., Nedorezov, T., Sharov, A., Zonderman, A. B., Abecasis, G. R., Costa, P., Lakatta, E., & Schlessinger, D. Heritability of cardiovascular and personality traits in 6,148 sardinians. *Plos Genetics* **2**, 1207-1223 (2006).

Polasek, O., Kolcic, I., Smoljanovic, A., Stojanovic, D., Grgic, M., Ebling, B., Klaric, M., Milas, J., & Puntaric, D. Demonstrating reduced environmental and genetic diversity in human isolates by analysis of blood lipid levels. *Croat.Med.J* **47**, 649-655 (2006).

Price N .C. (Ed) (1996). LABFAX Proteins. Bios Scientific Publishers.

Priller, C., Bauer, T., Mitteregger, G., Krebs, B., Kretzschmar, H. A., & Herms, J. Synapse formation and function is modulated by the amyloid precursor protein. *Journal of Neuroscience* **26**, 7212-7221 (2006).

Prince, J. A., Feuk, L., Gu, H. F., Johansson, B., Gatz, M., Blennow, K., & Brookes, A. J. Genetic variation in a haplotype block spanning IDE influences Alzheimer disease. *Hum.Mutat.* **22**, 363-371 (2003).

Prince, J. A., Feuk, L., Sawyer, S. L., Gottfries, J., Ricksten, A., Nagga, K., Bogdanovic, N., Blennow, K., & Brookes, A. J. Lack of replication of association findings in complex disease: an analysis of 15 polymorphisms in prior candidate genes for sporadic Alzheimer's disease. *European Journal of Human Genetics* **9**, 437-444 (2001).

Pritchard, J. K. & Rosenberg, N. A. Use of unlinked genetic markers to detect population stratification in association studies. *Am J Hum.Genet.* **65**, 220-228 (1999).

Probst, A., Gotz, J., Wiederhold, K. H., Tolnay, M., Mistl, C., Jaton, A. L., Hong, M., Ishihara, T., Lee, V. M., Trojanowski, J. Q., Jakes, R., Crowther, R. A., Spillantini, M. G., Burki, K., & Goedert, M. Axonopathy and amyotrophy in mice transgenic for human four-repeat tau protein. *Acta Neuropathol.* **99**, 469-481 (2000).

Prosser, B. E., Johnson, S., Roversi, P., Herbert, A. P., Blaum, B. S., Tyrrell, J., Jowitt, T. A., Clark, S. J., Tarelli, E., Uhrin, D., Barlow, P. N., Sim, R. B., Day, A. J., & Lea, S. M.

Structural basis for complement factor H linked age-related macular degeneration. *J Exp.Med.* **204**, 2277-2283 (2007).

Purcell, S., Cherny, S. S., & Sham, P. C. Genetic Power Calculator: design of linkage and association genetic mapping studies of complex traits. *Bioinformatics* **19**, 149-150 (2003).

Purcell, S., Neale, B., Todd-Brown, K., Thomas, L., Ferreira, M. A. R., Bender, D., Maller, J., Sklar, P., de Bakker, P. I. W., Daly, M. J., & Sham, P. C. PLINK: A tool set for whole-genome association and population-based linkage analyses. *American Journal of Human Genetics* **81**, 559-575 (2007).

Rajput-Williams, J., Knott, T. J., Wallis, S. C., Sweetnam, P., Yarnell, J., Cox, N., Bell, G. I., Miller, N. E., & Scott, J. Variation of apolipoprotein-B gene is associated with obesity, high blood cholesterol levels, and increased risk of coronary heart disease. *Lancet* **2**, 1442-1446 (1988).

Rall, S. C., Jr., Weisgraber, K. H., & Mahley, R. W. Human apolipoprotein E. The complete amino acid sequence. *J Biol.Chem.* **257**, 4171-4178 (1982).

Reich, D. E. & Lander, E. S. On the allelic spectrum of human disease. *Trends in Genetics* **17**, 502-510 (2001).

Richards, A., Buddles, M. R., Donne, R. L., Kaplan, B. S., Kirk, E., Venning, M. C., Tielemans, C. L., Goodship, J. A., & Goodship, T. H. Factor H mutations in hemolytic uremic syndrome cluster in exons 18-20, a domain important for host cell recognition. *Am J Hum.Genet.* **68**, 485-490 (2001).

Righetti, P. G., Castagna, A., Herbert, B., & Candiano, G. How to bring the "unseen" proteome to the limelight via electrophoretic pre-fractionation techniques. *Bioscience Reports* **25**, 3-17 (2005).

Ripoche, J., Day, A. J., Harris, T. J., & Sim, R. B. The complete amino acid sequence of human complement factor H. *Biochem.J* **249**, 593-602 (1988).

Risch, N. & Merikangas, K. The future of genetic studies of complex human diseases. *Science* **273**, 1516-1517 (1996).

Rivera, A., Fisher, S. A., Fritsche, L. G., Keilhauer, C. N., Lichtner, P., Meitinger, T., & Weber, B. H. F. Hypothetical LOC387715 is a second major susceptibility gene for age-related macular degeneration, contributing independently of complement factor H to disease risk. *Human Molecular Genetics* **14**, 3227-3236 (2005).

Rivera, A., White, K., Stohr, H., Steiner, K., Hemmrich, N., Grimm, T., Jurklies, B., Lorenz, B., Scholl, H. P., Apfelstedt-Sylla, E., & Weber, B. H. A comprehensive survey of sequence variation in the ABCA4 (ABCR) gene in Stargardt disease and age-related macular degeneration. *Am J Hum.Genet.* **67**, 800-813 (2000).

Robbins, R. B. Some Applications of Mathematics to Breeding Problems III. *Genetics* **3**, 375-389 (1918).

Ross, C. A. & Poirier, M. A. Protein aggregation and neurodegenerative disease. *Nature Medicine* S10-S17 (2004).

Ross, P. L., Huang, Y. L. N., Marchese, J. N., Williamson, B., Parker, K., Hattan, S., Khainovski, N., Pillai, S., Dey, S., Daniels, S., Purkayastha, S., Juhasz, P., Martin, S., Bartlet-Jones, M., He, F., Jacobson, A., & Pappin, D. J. Multiplexed protein quantitation in Saccharomyces cerevisiae using amine-reactive isobaric tagging reagents. *Molecular & Cellular Proteomics* **3**, 1154-1169 (2004).

Rozen, S. & Skaletsky, H. Primer3 on the WWW for general users and for biologist programmers. *Methods Mol.Biol.* **132**, 365-386 (2000).

Rubinsztein, D. C. & Easton, D. F. Apolipoprotein E genetic variation and Alzheimer's disease - A meta-analysis. *Dementia and Geriatric Cognitive Disorders* **10**, 199-209 (1999).

Rubinsztein, D. C. The genetics of Alzheimer's disease. *Progress in Neurobiology* **52**, 447-454 (1997).

Rudan, I. & Campbell, H. Five reasons why inbreeding may have considerable effect on post-reproductive human health. *Coll.Antropol.* **28**, 943-950 (2004).

Rudan, I. Inbreeding and cancer incidence in human isolates. *Hum.Biol.* **71**, 173-187 (1999a).

Rudan, I., Biloglav, Z., Vorko-Jovic, A., Kujundzic-Tiljak, M., Stevanovic, R., Ropac, D., Puntaric, D., Cucevic, B., Salzer, B., & Campbell, H. Effects of inbreeding, endogamy, genetic admixture, and outbreeding on human health: a (1001 Dalmatians) study. *Croat.Med.J* **47**, 601-610 (2006).

Rudan, I., Campbell, H., & Rudan, P. Genetic epidemiological studies of eastern Adriatic Island isolates, Croatia: objective and strategies. *Coll.Antropol.* **23**, 531-546 (1999b).

Rudan, P., Simic, D., Smolejnarancic, N., Bennett, L. A., Janicijevic, B., Jovanovic, V., Lethbridge, M. F., Milicic, J., Roberts, D. F., Sujoldzic, A., & Szirovicza, L. Isolation by Distance in Middle Dalmatia Yugoslavia. *American Journal of Physical Anthropology* **74**, 417-426 (1987).

Samani, N. J., Erdmann, J., Hall, A. S., Hengstenberg, C., Mangino, M., Mayer, B., Dixon, R. J., Meitinger, T., Braund, P., Wichmann, H. E., Barrett, J. H., Konig, I. R., Stevens, S. E., Szymczak, S., Tregouet, D. A., Iles, M. M., Pahlke, F., Pollard, H., Lieb, W., Cambien, F., Fischer, M., Ouwehand, W., Blankenberg, S., Balmforth, A. J., Baessler, A., Ball, S. G., Strom, T. M., Braenne, I., Gieger, C., Deloukas, P., Tobin, M. D., Ziegler, A., Thompson, J. R., & Schunkert, H. Genomewide association analysis of coronary artery disease. *N.Engl.J Med.* **357**, 443-453 (2007).

Sambrook J. & Russell D. W. (1989) Molecular Cloning, A Laboratory Manual. Cold Spring Harbor Laboratory Press.

Santangelo, S. L., Yen, C. H., Haddad, S., Fagerness, J., Huang, C., & Seddon, J. M. A discordant sib-pair linkage analysis of age-related macular degeneration. *Ophthalmic Genet.* **26**, 61-67 (2005).

Sassi, A. P., Andel, F., III, Bitter, H. M., Brown, M. P., Chapman, R. G., Espiritu, J., Greenquist, A. C., Guyon, I., Horchi-Alegre, M., Stults, K. L., Wainright, A., Heller, J. C., & Stults, J. T. An automated, sheathless capillary electrophoresis-mass spectrometry

platform for discovery of biomarkers in human serum. *Electrophoresis* **26**, 1500-1512 (2005).

Saunders, A. M., Strittmatter, W. J., Schmechel, D., Georgehyslop, P. H. S., Pericakvance, M. A., Joo, S. H., Rosi, B. L., Gusella, J. F., Crappermaclachlan, D. R., Alberts, M. J., Hulette, C., Crain, B., Goldgaber, D., & Roses, A. D. Association of Apolipoprotein-e Allele Epsilon-4 with Late-Onset Familial and Sporadic Alzheimers-Disease. *Neurology* **43**, 1467-1472 (1993).

Sayed-Tabatabaei, F. A., van Rijn, M. J., Schut, A. F., Aulchenko, Y. S., Croes, E. A., Zillikens, M. C., Pols, H. A., Witteman, J. C., Oostra, B. A., & van Duijn, C. M. Heritability of the function and structure of the arterial wall: findings of the Erasmus Rucphen Family (ERF) study. *Stroke* **36**, 2351-2356 (2005).

Schellenberg, G. D. Genetics of Alzheimer's disease. *Neurobiology of Aging* **25**, S1 (2004).

Scheuner, D., Eckman, C., Jensen, M., Song, X., Citron, M., Suzuki, N., Bird, T. D., Hardy, J., Hutton, M., Kukull, W., Larson, E., Levy-Lahad, E., Viitanen, M., Peskind, E., Poorkaj, P., Schellenberg, G., Tanzi, R., Wasco, W., Lannfelt, L., Selkoe, D., & Younkin, S. Secreted amyloid beta-protein similar to that in the senile plaques of Alzheimer's disease is increased in vivo by the presenilin 1 and 2 and APP mutations linked to familial Alzheimer's disease. *Nat.Med.* **2**, 864-870 (1996).

Schmidt, S., Hauser, M. A., Scott, W. K., Postel, E. A., Agarwal, A., Gallins, P., Wong, F., Chen, Y. S., Spencer, K., Schnetz-Boutaud, N., Haines, J. L., & Pericak-Vance, M. A. Cigarette smoking strongly modifies the association of LOC387715 and age-related macular degeneration. *Am J Hum.Genet.* **78**, 852-864 (2006).

Schmidt, S., Klaver, C., Saunders, A., Postel, E., De La, P. M., Agarwal, A., Small, K., Udar, N., Ong, J., Chalukya, M., Nesburn, A., Kenney, C., Domurath, R., Hogan, M., Mah, T., Conley, Y., Ferrell, R., Weeks, D., de Jong, P. T., van Duijn, C., Haines, J., Pericak-Vance, M., & Gorin, M. A pooled case-control study of the apolipoprotein E (APOE) gene in age-related maculopathy. *Ophthalmic Genet.* **23**, 209-223 (2002).

Schmidt, S., Postel, E. A., Agarwal, A., Allen, I. C., Jr., Walters, S. N., De La Paz, M. A., Scott, W. K., Haines, J. L., Pericak-Vance, M. A., & Gilbert, J. R. Detailed analysis of allelic variation in the ABCA4 gene in age-related maculopathy. *Invest Ophthalmol.Vis.Sci.* **44**, 2868-2875 (2003).

Schneider, W. J., Kovanen, P. T., Brown, M. S., Goldstein, J. L., Utermann, G., Weber, W., Havel, R. J., Kotite, L., Kane, J. P., Innerarity, T. L., & Mahley, R. W. Familial dysbetalipoproteinemia. Abnormal binding of mutant apoprotein E to low density lipoprotein receptors of human fibroblasts and membranes from liver and adrenal of rats, rabbits, and cows. *J Clin.Invest* **68**, 1075-1085 (1981).

Scuteri, A., Najjar, S. S., Muller, D., Andres, R., Morrell, C. H., Zonderman, A. B., & Lakatta, E. G. ApoE4 allele and the natural history of cardiovascular risk factors. *American Journal of Physiology-Endocrinology and Metabolism* **289**, E322-E327 (2005).

Seddon, J. M., Ajani, U. A., & Mitchell, B. D. Familial aggregation of age-related maculopathy. *Am J Ophthalmol.* **123**, 199-206 (1997).

Seddon, J. M., Ajani, U. A., Sperduto, R. D., Hiller, R., Blair, N., Burton, T. C., Farber, M. D., Gragoudas, E. S., Haller, J., Miller, D. T., & . Dietary carotenoids, vitamins A, C, and E, and advanced age-related macular degeneration. Eye Disease Case-Control Study Group. *JAMA* **272**, 1413-1420 (1994).

Seddon, J. M., Cote, J., & Rosner, B. Progression of age-related macular degeneration: association with dietary fat, transunsaturated fat, nuts, and fish intake. *Arch.Ophthalmol.* **121**, 1728-1737 (2003a).

Seddon, J. M., Cote, J., Davis, N., & Rosner, B. Progression of age-related macular degeneration: association with body mass index, waist circumference, and waist-hip ratio. *Arch.Ophthalmol.* **121**, 785-792 (2003b).

Seddon, J. M., Cote, J., Page, W. F., Aggen, S. H., & Neale, M. C. The US twin study of age-related macular degeneration: relative roles of genetic and environmental influences. *Arch.Ophthalmol.* **123**, 321-327 (2005).

Seddon, J. M., George, S., Rosner, B., & Klein, M. L. CFH gene variant, Y402H, and smoking, body mass index, environmental associations with advanced age-related macular degeneration. *Human Heredity* **61**, 157-165 (2006a).

Seddon, J. M., Santangelo, S. L., Book, K., Chong, S., & Cote, J. A genomewide scan for age-related macular degeneration provides evidence for linkage to several chromosomal regions. *Am J Hum.Genet.* **73**, 780-790 (2003c).

Seddon, J. M., Sharma, S., & Adelman, R. A. Evaluation of the clinical age-related maculopathy staging system. *Ophthalmology* **113**, 260-266 (2006b).

Seddon, J. M., Willett, W. C., Speizer, F. E., & Hankinson, S. E. A prospective study of cigarette smoking and age-related macular degeneration in women. *JAMA* **276**, 1141-1146 (1996).

Shadforth, I. P., Dunkley, T. P., Lilley, K. S., & Bessant, C. i-Tracker: for quantitative proteomics using iTRAQ. *BMC Genomics* **6**, 145 (2005).

Sham, P. C., Cherny, S. S., Purcell, S., & Hewitt, J. K. Power of linkage versus association analysis of quantitative traits, by use of variance-components models, for sibship data. *Am J Hum.Genet.* **66**, 1616-1630 (2000).

Shifman, S., Kuypers, J., Kokoris, M., Yakir, B., & Darvasi, A. Linkage disequilibrium patterns of the human genome across populations. *Hum.Mol.Genet.* **12**, 771-776 (2003).

Siegel, D. A., Davies, P., Dobrenis, K., & Huang, M. Tomoregulin-2 is found extensively in plaques in Alzheimer's disease brain. *J Neurochem.* **98**, 34-44 (2006).

Siepel, A., Bejerano, G., Pedersen, J. S., Hinrichs, A. S., Hou, M., Rosenbloom, K., Clawson, H., Spieth, J., Hillier, L. W., Richards, S., Weinstock, G. M., Wilson, R. K., Gibbs, R. A., Kent, W. J., Miller, W., & Haussler, D. Evolutionarily conserved elements in vertebrate, insect, worm, and yeast genomes. *Genome Res* **15**, 1034-1050 (2005).

Simonelli, F., Frisso, G., Testa, F., di Fiore, R., Vitale, D. F., Manitto, M. P., Brancato, R., Rinaldi, E., & Sacchetti, L. Polymorphism p.402Y>H in the complement factor H

protein is a risk factor for age related macular degeneration in an Italian population. *Br.J Ophthalmol.* **90**, 1142-1145 (2006).

Sing, C. F. & Davignon, J. Role of the apolipoprotein E polymorphism in determining normal plasma lipid and lipoprotein variation. *Am J Hum.Genet.* **37**, 268-285 (1985).

Skerka, C., Lauer, N., Weinberger, A. A., Keilhauer, C. N., Suhnel, J., Smith, R., Schlotzer-Schrehardt, U., Fritsche, L., Heinen, S., Hartmann, A., Weber, B. H., & Zipfel, P. F. Defective complement control of factor H (Y402H) and FHL-1 in age-related macular degeneration. *Mol.Immunol.* **44**, 3398-3406 (2007).

Smith, A. D., Johnston, C., Sim, E., Nagy, Z., Jobst, K. A., Hindley, N., & King, E. Protective Effect of Apoe Epsilon-2 in Alzheimers-Disease. *Lancet* **344**, 473-474 (1994).

Smith, D. J. & Lusis, A. J. The allelic structure of common disease. *Human Molecular Genetics* **11**, 2455-2461 (2002).

Smoljanovic, M., Ristic, S., & Hayward, C. Historic exposure to plague and present-day frequency of CCR5del32 in two isolated island communities of Dalmatia, Croatia. *Croat.Med.J* **47**, 579-584 (2006).

Snow, A. D., Kinsella, M. G., Parks, E., Sekiguchi, R. T., Miller, J. D., Kimata, K., & Wight, T. N. Differential binding of vascular cell-derived proteoglycans (perlecan, biglycan, decorin, and versican) to the beta-amyloid protein of Alzheimer's disease. *Arch.Biochem.Biophys.* **320**, 84-95 (1995).

Sokal R. R. & Rohlf F. J. (2000). Biometry (2nd Edition). W H Freeman & Co, San Francisco.

Souied, E. H., Benlian, P., Amouyel, P., Feingold, J., Lagarde, J. P., Munnich, A., Kaplan, J., Coscas, G., & Soubrane, G. The epsilon 4 allele of the apolipoprotein E gene as a potential protective factor for exudative age-related macular degeneration. *American Journal of Ophthalmology* **125**, 353-359 (1998).

Spector, D. L., Ochs, R. L., & Busch, H. Silver staining, immunofluorescence, and immunoelectron microscopic localization of nucleolar phosphoproteins B23 and C23. *Chromosoma* **90**, 139-148 (1984).

Spencer, K. L., Hauser, M. A., Olson, L. M., Schmidt, S., Scott, W. K., Gallins, P., Agarwal, A., Postel, E. A., Pericak-Vance, M. A., & Haines, J. L. Protective effect of complement factor B and complement component 2 variants in age-related macular degeneration. *Hum.Mol.Genet.* **16**, 1986-1992 (2007a).

Spencer, K. L., Hauser, M. A., Olson, L. M., Schnetz-Boutaud, N., Scott, W. K., Schmidt, S., Gallins, P., Agarwal, A., Postel, E. A., Pericak-Vance, M. A., & Haines, J. L. Haplotypes spanning the complement factor H gene are protective against age-related macular degeneration. *Investigative Ophthalmology & Visual Science* **48**, 4277-4283 (2007b).

Spires, T. L. & Hyman, B. T. Neuronal structure is altered by amyloid plaques. *Reviews in the Neurosciences* **15**, 267-278 (2004).

Spires, T. L. & Hyman, B. T. Transgenic models of Alzheimer's disease: learning from animals. *NeuroRx.* **2**, 423-437 (2005).

Stacey, S. N., Manolescu, A., Sulem, P., Rafnar, T., Gudmundsson, J., Gudjonsson, S. A., Masson, G., Jakobsdottir, M., Thorlacius, S., Helgason, A., Aben, K. K., Strobbe, L. J., Albers-Akkers, M. T., Swinkels, D. W., Henderson, B. E., Kolonel, L. N., Le Marchand, L., Millastre, E., Andres, R., Godino, J., Garcia-Prats, M. D., Polo, E., Tres, A., Mouy, M., Saemundsdottir, J., Backman, V. M., Gudmundsson, L., Kristjansson, K., Bergthorsson, J. T., Kostic, J., Frigge, M. L., Geller, F., Gudbjartsson, D., Sigurdsson, H., Jonsdottir, T., Hrafnkelsson, J., Johannsson, J., Sveinsson, T., Myrdal, G., Grimsson, H. N., Jonsson, T., von Holst, S., Werelius, B., Margolin, S., Lindblom, A., Mayordomo, J. I., Haiman, C. A., Kiemeney, L. A., Johannsson, O. T., Gulcher, J. R., Thorsteinsdottir, U., Kong, A., & Stefansson, K. Common variants on chromosomes 2q35 and 16q12 confer susceptibility to estrogen receptor-positive breast cancer. *Nature Genetics* **39**, 865-869 (2007).

Stacey, S. N., Sulem, P., Johannsson, O. T., Helgason, A., Gudmundsson, J., Kostic, J. P., Kristjansson, K., Jonsdottir, T., Sigurdsson, H., Hrafnkelsson, J., Johannsson, J., Sveinsson, T., Myrdal, G., Grimsson, H. N., Bergthorsson, J. T., Amundadottir, L. T., Gulcher, J. R., Thorsteinsdottir, U., Kong, A., & Stefansson, K. The BARD1 Cys557Ser variant and breast cancer risk in Iceland. *Plos Medicine* **3**, 1103-1113 (2006).

Stefansson, H., Rye, D. B., Hicks, A., Petursson, H., Ingason, A., Thorgeirsson, T. E., Palsson, S., Sigmundsson, T., Sigurdsson, A. P., Eiriksdottir, I., Soebech, E., Bliwise, D., Beck, J. M., Rosen, A., Waddy, S., Trotti, L. M., Iranzo, A., Thambisetty, M., Hardarson, G. A., Kristjansson, K., Gudmundsson, L. J., Thorsteinsdottir, U., Kong, A., Gulcher, J. R., Gudbjartsson, D., & Stefansson, K. A genetic risk factor for periodic limb movements in sleep. *New England Journal of Medicine* **357** , 639-647 (2007).

Stefansson, S. E., Jonsson, H., Ingvarsson, T., Manolescu, I., Jonsson, H. H., Olafsdottir, G., Palsdottir, E., Stefansdottir, G., Sveinbjornsdottir, G., Frigge, M. L., Kong, A., Gulcher, J. R., & Stefansson, K. Genomewide scan for hand osteoarthritis: A novel mutation in matrilin-3. *American Journal of Human Genetics* **72**, 1448-1459 (2003).

Stgeorgehyslop, P. H., Tanzi, R. E., Polinsky, R. J., Haines, J. L., Nee, L., Watkins, P. C., Myers, R. H., Feldman, R. G., Pollen, D., Drachman, D., Growdon, J., Bruni, A., Foncin, J. F., Salmon, D., Frommelt, P., Amaducci, L., Sorbi, S., Piacentini, S., Stewart, G. D., Hobbs, W. J., Conneally, P. M., & Gusella, J. F. The Genetic-Defect Causing Familial Alzheimers-Disease Maps on Chromosome-21. *Science* **235**, 885-890 (1987).

Stone, E. M., Braun, T. A., Russell, S. R., Kuehn, M. H., Lotery, A. J., Moore, P. A., Eastman, C. G., Casavant, T. L., & Sheffield, V. C. Missense variations in the fibulin 5 gene and age-related macular degeneration. *New England Journal of Medicine* **351** , 346-353 (2004).

Stone, E. M., Webster, A. R., Vandenburgh, K., Streb, L. M., Hockey, R. R., Lotery, A. J., & Sheffield, V. C. Allelic variation in ABCR associated with Stargardt disease but not age-related macular degeneration. *Nat.Genet.* **20**, 328-329 (1998).

Storey, J. D. & Tibshirani, R. Statistical significance for genomewide studies. *Proc.Natl.Acad.Sci.U.S A* **100**, 9440-9445 (2003).

Stratman, N. C., Castle, C. K., Taylor, B. M., Epps, D. E., Melchior, G. W., & Carter, D. B. Isoform-specific interactions of human apolipoprotein E to an intermediate conformation of human Alzheimer amyloid-beta peptide. *Chemistry and Physics of Lipids* **137** , 52-61 (2005).

Strittmatter, W. J., Weisgraber, K. H., Huang, D. Y., Dong, L. M., Salvesen, G. S., Pericak-Vance, M., Schmechel, D., Saunders, A. M., Goldgaber, D., & Roses, A. D. Binding of human apolipoprotein E to synthetic amyloid beta peptide: isoform-specific effects and implications for late-onset Alzheimer disease. *Proc.Natl.Acad.Sci.U.S A* **90**, 8098-8102 (1993).

Sturchler-Pierrat, C., Abramowski, D., Duke, M., Wiederhold, K. H., Mistl, C., Rothacher, S., Ledermann, B., Burki, K., Frey, P., Paganetti, P. A., Waridel, C., Calhoun, M. E., Jucker, M., Probst, A., Staufenbiel, M., & Sommer, B. Two amyloid precursor protein transgenic mouse models with Alzheimer disease-like pathology. *Proc.Natl.Acad.Sci.U.S A* **94**, 13287-13292 (1997).

Sujoldzic, A., Jovanovic, V., Angel, J. L., Bennett, L. A., Roberts, D. F., & Rudan, P. Migration within the island of Korcula, Yugoslavia. *Ann.Hum.Biol.* **16**, 483-493 (1989).

Suzuki, N., Cheung, T. T., Cai, X. D., Odaka, A., Otvos, L., Jr., Eckman, C., Golde, T. E., & Younkin, S. G. An increased percentage of long amyloid beta protein secreted by familial amyloid beta protein precursor (beta APP717) mutants. *Science* **264**, 1336-1340 (1994).

Swaroop, A., Branham, K. E. H., Chen, W., & Abecasis, G. Genetic susceptibility to age-related macular degeneration: a paradigm for dissecting complex disease traits. *Human Molecular Genetics* **16**, R174-R182 (2007).

Switzer, R. C., III, Merril, C. R., & Shifrin, S. A highly sensitive silver stain for detecting proteins and peptides in polyacrylamide gels. *Anal.Biochem.* **98**, 231-237 (1979).

Tatusova, T. A. & Madden, T. L. BLAST 2 Sequences, a new tool for comparing protein and nucleotide sequences. *FEMS Microbiol.Lett.* **174**, 247-250 (1999).

Taylor J. K. (1990). Statistical techniques for Data Analysis. Lewis Publishers, Chelsea, Michigan.

Terai, K., Iwai, A., Kawabata, S., Tasaki, Y., Watanabe, T., Miyata, K., & Yamaguchi, T. beta-amyloid deposits in transgenic mice expressing human beta-amyloid precursor protein have the same characteristics as those in Alzheimer's disease. *Neuroscience* **104**, 299-310 (2001).

Terwilliger, J. D. & Weiss, K. M. Linkage disequilibrium mapping of complex disease: fantasy or reality? *Curr.Opin.Biotechnol.* **9**, 578-594 (1998).

Thomas, P. D. & Kejariwal, A. Coding single-nucleotide polymorphisms associated with complex vs. Mendelian disease: evolutionary evidence for differences in molecular effects. *Proc.Natl.Acad.Sci.U.S A* **101**, 15398-15403 (2004).

Thorleifsson, G., Magnusson, K. P., Sulem, P., Walters, G. B., Gudbjartsson, D. F., Stefansson, H., Jonsson, T., Jonasdottir, A., Jonasdottir, A., Stefansdottir, G., Masson, G., Hardarson, G. A., Petursson, H., Arnarsson, A., Motallebipour, M., Wallerman, O.,

Wadelius, C., Gulcher, J. R., Thorsteinsdottir, U., Kong, A., Jonasson, F., & Stefansson, K. Common sequence variants in the LOXL1 gene confer susceptibility to exfoliation glaucoma. *Science* **317**, 1397-1400 (2007).

Thornton, J., Edwards, R., Mitchell, P., Harrison, R. A., Buchan, I., & Kelly, S. P. Smoking and age-related macular degeneration: a review of association. *Eye* **19**, 935-944 (2005).

Tirumalai, R. S., Chan, K. C., Prieto, D. A., Issaq, H. J., Conrads, T. P., & Veenstra, T. D. Characterization of the low molecular weight human serum proteome. *Molecular & Cellular Proteomics* **2**, 1096-1103 (2003).

Todd, J. A., Bell, J. I., & McDevitt, H. O. HLA-DQ beta gene contributes to susceptibility and resistance to insulin-dependent diabetes mellitus. *Nature* **329**, 599-604 (1987).

Uchida, T., Fukawa, A., Uchida, M., Fujita, K., & Saito, K. Application of a novel protein biochip technology for detection and identification of rheumatoid arthritis biomarkers in synovial fluid. *Journal of Proteome Research* **1**, 495-499 (2002).

Uka, J., Tamura, H., Kobayashi, T., Yamane, K., Kawakami, H., Minamoto, A., & Mishima, H. K. No association of complement factor H gene polymorphism and age-related macular degeneration in the Japanese population. *Retina* **26**, 985-987 (2006).

Utermann, G. Apolipoprotein E mutants, hyperlipidemia and arteriosclerosis. *Adv.Exp.Med.Biol.* **183**, 173-188 (1985).

Utermann, G., Pruin, N., & Steinmetz, A. Polymorphism of apolipoprotein E. III. Effect of a single polymorphic gene locus on plasma lipid levels in man. *Clin.Genet.* **15**, 63-72 (1979).

van der Werff, M. P. J., Mertens, B., de Noo, M. E., Bladergroen, M. R., Dalebout, H. C., Tollenaar, R. A. E. M., & Deelder, A. M. Case-control breast cancer study of MALDI-TOF proteomic mass spectrometry data on serum samples. *Statistical Applications in Genetics and Molecular Biology* **7**, (2008).

van Duijn, C. M., de Knijff, P., Cruts, M., Wehnert, A., Havekes, L. M., Hofman, A., & van Broeckhoven, C. Apolipoprotein E4 allele in a population-based study of early-onset Alzheimer's disease. *Nat.Genet.* **7**, 74-78 (1994).

van Oijen, M., Hofman, A., Soares, H. D., Koudstaal, P. J., & Breteler, M. M. Plasma Abeta(1-40) and Abeta(1-42) and the risk of dementia: a prospective case-cohort study. *Lancet Neurol.* **5**, 655-660 (2006).

Van Weemen BK & Schuurs, A. H. W. Immunoassay Using Antigen-Enzyme Conjugates. *Febs Letters* **15**, 232-& (1971).

VanBroeckhoven, C., Haan, J., Bakker, E., Hardy, J. A., Vanhul, W., Wehnert, A., Vegtervandervlis, M., & Roos, R. A. C. Amyloid-Beta Protein-Precursor Gene and Hereditary Cerebral-Hemorrhage with Amyloidosis (Dutch). *Science* **248**, 1120-1122 (1990).

Vetrivel, K. S. & Thinakaran, G. Amyloidogenic processing of beta-amyloid precursor protein in intracellular compartments. *Neurology* **66**, S69-S73 (2006).

Villanueva, J., Philip, J., Entenberg, D., Chaparro, C. A., Tanwar, M. K., Holland, E. C., & Tempst, P. Serum peptide profiling by magnetic particle-assisted, automated sample processing and MALDI-TOF mass spectrometry. *Analytical Chemistry* **76**, 1560-1570 (2004).

Villanueva, J., Shaffer, D. R., Philip, J., Chaparro, C. A., Erdjument-Bromage, H., Olshen, A. B., Fleisher, M., Lilja, H., Brogi, E., Boyd, J., Sanchez-Carbayo, M., Holland, E. C., Cordon-Cardo, C., Scher, H. I., & Tempst, P. Differential exoprotease activities confer tumor-specific serum peptidome patterns. *Journal of Clinical Investigation* **116**, 271-284 (2006).

Vingerling, J. R., Klaver, C. C., Hofman, A., & de Jong, P. T. Epidemiology of age-related maculopathy. *Epidemiol.Rev.* **17**, 347-360 (1995).

Vitart, V., Biloglav, Z., Hayward, C., Janicijevic, B., Smolej-Narancic, N., Barac, L., Pericic, M., Klaric, I. M., Skaric-Juric, T., Barbalic, M., Polasek, O., Kolcic, I., Carothers, A., Rudan, P., Hastie, N., Wright, A., Campbell, H., & Rudan, I. 3000 years of solitude: extreme differentiation in the island isolates of Dalmatia, Croatia. *European Journal of Human Genetics* **14**, 478-487 (2006).

Vitart, V., Carothers, A. D., Hayward, C., Teague, P., Hastie, N. D., Campbell, H., & Wright, A. F. Increased level of linkage disequilibrium in rural compared with urban communities: A factor to consider in association-study design. *American Journal of Human Genetics* **76**, 763-772 (2005).

Vitart, V., Rudan, I., Hayward, C., Gray, N. K., Floyd, J., Palmer, C. N., Knott, S. A., Kolcic, I., Polasek, O., Graessler, J., Wilson, J. F., Marinaki, A., Riches, P. L., Shu, X., Janicijevic, B., Smolej-Narancic, N., Gorgoni, B., Morgan, J., Campbell, S., Biloglav, Z., Barac-Lauc, L., Pericic, M., Klaric, I. M., Zgaga, L., Skaric-Juric, T., Wild, S. H., Richardson, W. A., Hohenstein, P., Kimber, C. H., Tenesa, A., Donnelly, L. A., Fairbanks, L. D., Aringer, M., McKeigue, P. M., Ralston, S. H., Morris, A. D., Rudan, P., Hastie, N. D., Campbell, H., & Wright, A. F. SLC2A9 is a newly identified urate transporter influencing serum urate concentration, urate excretion and gout. *Nat.Genet.* **40**, 437-442 (2008).

Vlachantoni D. (2006). The Role of Oxidative Stress on Photoreceptor Degeneration (Thesis). University of Edinburgh.

Voight, B. F., Kudaravalli, S., Wen, X., & Pritchard, J. K. A map of recent positive selection in the human genome. *PLoS Biol.* **4**, e72 (2006).

Vuletic, M. G. & Mujkic, A. Mental health and health-related quality of life in Croatian island population. *Croat.Med.J* **47**, 635-640 (2006).

Wacholder, S., Chanock, S., Garcia-Closas, M., El Ghormli, L., & Rothman, N. Assessing the probability that a positive report is false: an approach for molecular epidemiology studies. *J Natl.Cancer Inst.* **96**, 434-442 (2004).

Waddle, D. M., Sokal, R. R., & Rudan, P. Factors affecting population variation in eastern Adriatic isolates (Croatia). *Hum.Biol.* **70**, 845-864 (1998).

Walker, E. S., Martinez, M., Brunkan, A. L., & Goate, A. Presenilin 2 familial Alzheimer's disease mutations result in partial loss of function and dramatic changes in A beta 42/40 ratios. *Journal of Neurochemistry* **92**, 294-301 (2005).

Wall, J. D. & Pritchard, J. K. Haplotype blocks and linkage disequilibrium in the human genome. *Nat.Rev.Genet.* **4**, 587-597 (2003).

Walport, M. J. Complement. First of two parts. *N.Engl.J Med.* **344**, 1058-1066 (2001).

Wang, J., Dickson, D. W., Trojanowski, J. Q., & Lee, V. M. Y. The levels of soluble versus insoluble brain A beta distinguish Alzheimer's disease from normal and pathologic aging. *Experimental Neurology* **158**, 328-337 (1999).

Wang, W. Y. S., Barratt, B. J., Clayton, D. G., & Todd, J. A. Genome-wide association studies: Theoretical and practical concerns. *Nature Reviews Genetics* **6**, 109-118 (2005).

Warwicker, P., Goodship, T. H., Donne, R. L., Pirson, Y., Nicholls, A., Ward, R. M., Turnpenny, P., & Goodship, J. A. Genetic studies into inherited and sporadic hemolytic uremic syndrome. *Kidney Int.* **53**, 836-844 (1998).

Weeks, D. E., Conley, Y. P., Tsai, H. J., Mah, T. S., Schmidt, S., Postel, E. A., Agarwal, A., Haines, J. L., Pericak-Vance, M. A., Rosenfeld, P. J., Paul, T. O., Eller, A. W., Morse, L. S., Dailey, J. P., Ferrell, R. E., & Gorin, M. B. Age-related maculopathy: A genomewide scan with continued evidence of susceptibility loci within the 1q31, 10q26, and 17q25 regions. *American Journal of Human Genetics* **75**, 174-189 (2004).

Weisgraber, K. H., Innerarity, T. L., & Mahley, R. W. Abnormal lipoprotein receptor-binding activity of the human E apoprotein due to cysteine-arginine interchange at a single site. *J Biol.Chem.* **257**, 2518-2521 (1982).

Weisgraber, K. H., Rall, S. C., Jr., & Mahley, R. W. Human E apoprotein heterogeneity. Cysteine-arginine interchanges in the amino acid sequence of the apo-E isoforms. *J Biol.Chem.* **256**, 9077-9083 (1981).

Weissinger, E. M., Wittke, S., Kaiser, T., Haller, H., Bartel, S., Krebs, R., Golovko, I., Rupprecht, H. D., Haubitz, M., Hecker, H., Mischak, H., & Fliser, D. Proteomic patterns established with capillary electrophoresis and mass spectrometry for diagnostic purposes. *Kidney Int.* **65**, 2426-2434 (2004).

Wellcome Trust Case Control Consortium. Genome-wide association study of 14,000 cases of seven common diseases and 3,000 shared controls. *Nature* **447**, 661-678 (2007).

Wenham, P. R., Price, W. H., & Blandell, G. Apolipoprotein E genotyping by one-stage PCR. *Lancet* **337**, 1158-1159 (1991).

Whaley, K. & Ruddy, S. Modulation of C3b hemolytic activity by a plasma protein distinct from C3b inactivator. *Science* **193**, 1011-1013 (1976).

Wigginton, J. E., Cutler, D. J., & Abecasis, G. R. A note on exact tests of Hardy-Weinberg equilibrium. *Am J Hum.Genet.* **76**, 887-893 (2005).

Williams, N. M., Norton, N., Williams, H., Ekholm, B., Hamshere, M. L., Lindblom, Y., Chowdari, K. V., Cardno, A. G., Zammit, S., Jones, L. A., Murphy, K. C., Sanders, R.

D., McCarthy, G., Gray, M. Y., Jones, G., Holmans, P., Nimgaonkar, V., Adolfson, R., Osby, U., Terenius, L., Sedvall, G., O'Donovan, M. C., & Owen, M. J. A systematic genomewide linkage study in 353 sib pairs with schizophrenia. *Am J Hum.Genet.* **73** , 1355-1367 (2003).

Wiszniewski, W., Zaremba, C. M., Yatsenko, A. N., Jamrich, M., Wensel, T. G., Lewis, R. A., & Lupski, J. R. ABCA4 mutations causing mislocalization are found frequently in patients with severe retinal dystrophies. *Human Molecular Genetics* **14**, 2769-2778 (2005).

Wood, W. G., Schroeder, F., Igbavboa, U., Avdulov, N. A., & Chochina, S. V. Brain membrane cholesterol domains, aging and amyloid beta-peptides. *Neurobiol.Aging* **23**, 685-694 (2002).

Workman, P. L., Lucarelli, P., Agostino, R., Scarabino, R., Scacchi, R., Carapella, E., Palmarino, R., & Bottini, E. Genetic differentiation among Sardinian villages. *Am J Phys.Anthropol.* **43**, 165-176 (1975).

Wright, A. F., Carothers, A. D., & Pirastu, M. Population choice in mapping genes for complex diseases. *Nature Genetics* **23**, 397-404 (1999).

Wu, S. L., Choudhary, G., Ramstrom, M., Bergquist, J., & Hancock, W. S. Evaluation of shotgun sequencing for proteomic analysis of human plasma using HPLC coupled with either ion trap or Fourier transform mass spectrometry. *Journal of Proteome Research* **2**, 383-393 (2003).

Yamazaki, T., Haass, C., Saido, T. C., Omura, S., & Ihara, Y. Specific increase in amyloid beta-protein 42 secretion ratio by calpain inhibition. *Biochemistry* **36**, 8377-8383 (1997).

Yan, Q., Zhang, J., Liu, H., Babu-Khan, S., Vassar, R., Biere, A. L., Citron, M., & Landreth, G. Anti-inflammatory drug therapy alters beta-amyloid processing and deposition in an animal model of Alzheimer's disease. *J Neurosci.* **23**, 7504-7509 (2003).

Yan, W., Lee, H., Deutsch, E. W., Lazaro, C. A., Tang, W. L., Chen, E., Fausto, N., Katze, M. G., & Aebersold, R. A dataset of human liver proteins identified by protein profiling via isotope-coded affinity tag (ICAT) and tandem mass spectrometry. *Molecular & Cellular Proteomics* **3**, 1039-1041 (2004).

Yang, Z. L., Camp, N. J., Sun, H., Tong, Z. Z., Gibbs, D., Cameron, D. J., Chen, H. Y., Zhao, Y., Pearson, E., Li, X., Chien, J., Dewan, A., Harmon, J., Bernstein, P. S., Shridhar, V., Zabriskie, N. A., Hoh, J., Howes, K., & Zhang, K. A variant of the HTRA1 gene increases susceptibility to age-related macular degeneration. *Science* **314**, 992-993 (2006).

Yates, J. R. W., Sepp, T., Matharu, B. K., Khan, J. C., Thurlby, D. A., Shahid, H., Clayton, D. G., Hayward, C., Morgan, J., Wright, A. F., Armbrecht, A. M., Dhillon, B., Deary, I. J., Redmond, E., Bird, A. C., & Moore, A. T. Complement C3 variant and the risk of age-related macular degeneration. *New England Journal of Medicine* **357**, 553-561 (2007).

Yocum, A. K., Yu, K., Oe, T., & Blair, I. A. Effect of immunoaffinity depletion of human serum during proteomic investigations. *J Proteome Res* **4**, 1722-1731 (2005).

Yoshida, T., Dewan, A., Zhang, H., Sakamoto, R., Okamoto, H., Minami, M., Obazawa, M., Mizota, A., Tanaka, M., Saito, Y., Takagi, I., Hoh, J., & Iwata, T. HTRA1 promoter polymorphism predisposes Japanese to age-related macular degeneration. *Mol.Vis.* **13**, 545-548 (2007).

Yoshizawa, M., Kawauchi, T., Sone, M., Nishimura, Y. V., Terao, M., Chihama, K., Nabeshima, Y., & Hoshino, M. Involvement of a Rac activator,P-Rex1, in neurotrophin-derived signaling and neuronal migration. *J Neurosci.* **25**, 4406-4419 (2005).

Younkin, S. G. The role of A beta 42 in Alzheimer's disease. *J Physiol Paris* **92**, 289-292 (1998).

Yu, J. M., Wiita, P., Kawaguchi, R., Honda, J., Jorgensen, A., Zhang, K., Fischetti, V. A., & Sun, H. Biochemical analysis of a common human polymorphism associated with age-related macular degeneration. *Biochemistry* **46**, 8451-8461 (2007).

Yu, J., Wiita, P., Kawaguchi, R., Honda, J., Jorgensen, A., Zhang, K., Fischetti, V. A., & Sun, H. Biochemical analysis of a common human polymorphism associated with age-related macular degeneration. *Biochemistry* **46**, 8451-8461 (2007).

Zamani, M. R. & Allen, Y. S. Nicotine and its interaction with beta-amyloid protein: a short review. *Biol.Psychiatry* **49**, 221-232 (2001).

Zareparsi, S., Branham, K. E. H., Li, M. Y., Shah, S., Klein, R. J., Ott, J., Hoh, J., Abecasis, G. R., & Swaroop, A. Strong association of the Y402H variant in complement factor H at 1q32 with susceptibility to age-related macular degeneration. *American Journal of Human Genetics* **77**, 149-153 (2005a).

Zareparsi, S., Buraczynska, M., Branham, K. E., Shah, S., Eng, D., Li, M., Pawar, H., Yashar, B. M., Moroi, S. E., Lichter, P. R., Petty, H. R., Richards, J. E., Abecasis, G. R., Elner, V. M., & Swaroop, A. Toll-like receptor 4 variant D299G is associated with susceptibility to age-related macular degeneration. *Hum.Mol.Genet.* **14**, 1449-1455 (2005b).

Zarghooni, M., Soosaipillai, A., Grass, L., Scorilas, A., Mirazimi, N., & Diamandis, E. P. Decreased concentration of human kallikrein 6 in brain extracts of Alzheimer's disease patients. *Clin.Biochem.* **35**, 225-231 (2002).

Zhang, R., Barker, L., Pinchev, D., Marshall, J., Rasamoelisolo, M., Smith, C., Kupchak, P., Kireeva, I., Ingratta, L., & Jackowski, G. Mining biomarkers in human sera using proteomic tools. *Proteomics* **4**, 244-256 (2004a).

Zhang, X., Leung, S. M., Morris, C. R., & Shigenaga, M. K. Evaluation of a novel, integrated approach using functionalized magnetic beads, bench-top MALDI-TOF-MS with prestructured sample supports, and pattern recognition software for profiling potential biomarkers in human plasma. *J Biomol.Tech.* **15**, 167-175 (2004b).

Zhang, Z. H., Hartmann, H., Do, V. M., Abramowski, D., Sturchler-Pierrat, C., Staufenbiel, M., Sommer, B., van de Wetering, M., Clevers, H., Saftig, P., De Strooper, B., He, X., & Yankner, B. A. Destabilization of beta-catenin by mutations in presenilin-1 potentiates neuronal apoptosis. *Nature* **395**, 698-702 (1998).

Zhao, G. J., Cui, M. Z., Mao, G. Z., Dong, Y. Z., Tan, J. X., Sun, L. S., & Xu, X. M. gamma-cleavage is dependent on zeta-cleavage during the proteolytic processing of amyloid precursor protein within its transmembrane domain. *Journal of Biological Chemistry* **280**, 37689-37697 (2005).

Zhao, G. J., Mao, G. Z., Tan, J. X., Dong, Y. Z., Cui, M. Z., Kim, S. H., & Xu, X. M. Identification of a new presenilin-dependent zeta-cleavage site within the transmembrane domain of amyloid precursor protein. *Journal of Biological Chemistry* **279**, 50647-50650 (2004).

Zheng, C., Heintz, N., & Hatten, M. E. CNS gene encoding astrotactin, which supports neuronal migration along glial fibers. *Science* **272**, 417-419 (1996).

Ziegler, A., Konig, I. R., & Thompson, J. R. Biostatistical aspects of genome-wide association studies. *Biom.J* **50**, 8-28 (2008).

Zinellu, A., Caria, M. A., Tavera, C., Sotgia, S., Chessa, R., Delana, L., & Carru, C. Plasma creatinine and creatine quantification by capillary electrophoresis diode array detector. *Analytical Biochemistry* **342**, 186-193 (2005).

Zipfel, P. F. & Skerka, C. Complement factor H and related proteins: an expanding family of complement-regulatory proteins? *Immunol.Today* **15**, 121-126 (1994).

Zipfel, P. F. Complement factor H: Physiology and pathophysiology. *Seminars in Thrombosis and Hemostasis* **27**, 191-199 (2001).

Zipfel, P. F., Heinen, S., Jozsi, M., & Skerka, C. Complement and diseases: Defective alternative pathway control results in kidney and eye diseases. *Molecular Immunology* **43**, 97-106 (2006).

Zlokovic, B. V. Clearing amyloid through the blood-brain barrier. *J Neurochem.* **89**, 807-811 (2004).

Zou, G. Y. & Donner, A. The merits of testing Hardy-Weinberg equilibrium in the analysis of unmatched case-control data: a cautionary note. *Ann.Hum.Genet.* **70**, 923-933 (2006).

Zurdel, J., Finckh, U., Menzer, G., Nitsch, R. M., & Richard, G. CST3 genotype associated with exudative age related macular degeneration. *Br.J Ophthalmol.* **86**, 214-219 (2002).